

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 04:53:07 : Search time 217.98 Seconds  
(without alignments)  
8664.125 Million cell updates/sec

Title: US-09-896-791B-2

Perfect score: 1100  
Sequence: 1 gaattcgacagagggccat.....aaaaaaaaacatgagcgc 1100

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	556.6	50.6	2207	20	AAK58986
2	471.6	42.9	2082	22	AAH14706
3	447.4	40.7	2595	22	AAH17091
4	418.8	38.1	711	22	AAH07034
5	392.8	35.7	756	22	AAH03708
6	253.4	23.0	1708	22	AAH15203
7	253.4	23.0	873	22	AAH03186
8	211.6	19.2	733	22	AAH03810
9	211.6	19.2	1841	22	AAH18364

10	174.6	15.9	2730	20	AAK58981	Human transcriptio
11	173	15.7	1607	21	AAZ94052	Human endothelial
12	173	15.7	2816	19	AAV00541	CDNA encoding huma
13	173	15.7	2818	21	AAZ94053	DNA encoding huma
14	173	15.7	32169	22	ABA14358	Human nervous syst
15	172	15.6	2776	22	AAH14355	Human EPAS1 DNA us
16	156.6	14.2	3031	19	AAV00642	CDNA encoding mur
17	138.2	12.6	3746	24	AAV09710	Mouse ischaemic m
18	134.6	12.2	2588	22	AAH14154	Human HIF-1 alpha
19	134.6	12.2	3678	18	AAV63210	Nucleic acid sequ
20	134.6	12.2	3736	18	AAV45937	Human hypoxia indu
21	134.6	12.2	3736	21	AAZ99537	DNA encoding a w
22	134.6	12.2	3927	24	AAH61690	Lung small cell ca
23	134.6	12.2	3933	20	AAK58980	Human transcriptio
24	134.6	12.2	10355	22	AAH14156	PSMAK7 CMT-HIF DNA
25	102.2	9.3	4162	23	AAH85058	DNA encoding novel
26	97.4	8.9	2487	23	AAH85055	DNA encoding novel
27	92	8.4	139	21	AAV41256	Human secreted pro
28	89.6	8.1	2078	19	AAV41256	Human neuronal PAS
29	83.2	7.6	4565	21	AAZ33295	Drosophila melanog
30	83.2	7.6	4571	23	ABH11895	Drosophila melanog
31	83.2	7.6	12776	23	ABH11894	Drosophila melanog
32	79	7.2	1475	21	AAH18197	Lung cancer associ
33	72.6	6.6	3614	18	AAH91883	Murine SIM (single
34	72.2	6.6	238	18	AAH91880	Partial human SIM
35	72	6.5	2082	19	AAV41257	Mouse neuronal PAS
36	69.2	6.3	549	21	AAZ93991	Cat flea hindgut a
37	68.4	6.2	861	21	AAH40542	Human fetal brain
38	67.6	6.1	561	22	AAH09235	Human CDNA clone (
39	66.4	6.0	27884	22	AAK77781	Human immune/haema
40	65.8	6.0	411	22	AAH93716	Human polynucleoti
41	65.6	6.0	416	22	AAH8617	Human polynucleoti
42	65.6	6.0	1432	24	ABH34259	Human immune syste
43	65.2	5.9	449	22	AAH16186	Human breast cance
44	65.2	5.9	1835	21	AAH18013	Lung cancer associ
45	65.2	5.9	2058	22	AAH26006	Human CDNA encodin

## ALIGNMENTS

RESULT 1	
ID	AAK58986 standard; CDNA; 2207 BP.
AC	AAK58986;
XX	
AC	23-AUG-1999 (first entry)
DT	
XX	
DE	Mouse transcription regulator MOP7 CDNA.
XX	
KW	MOP7, member of the PAS superfamily; bHLH-PAS; mouse;
KW	transcription regulator; hypoxia inducible factor 3 alpha; ss.
XX	
OS	Mus musculus.
XX	
FH	
FT	Key
FT	CDS
XX	
PN	W0928464-A2.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98MO-US25314.
XX	
PR	28-NOV-1997; 97US-0066863.
XX	
PA	(WISC ) WISCONSIN ALUMNI RES FOUND.
XX	
PI	Bradfield CA, Gu YZ, Hogenesch JB;
XX	
DR	WPI; 1999-371120/31.





```
QY 224 tctgagc-----agcagctgtaaaagggagagccactgagcgcctgct 269
DB 263 tctgagcagggagagtggaacacaggtgggagcagggaggaacacagctgctgct 322
QY 270 acctgaagggcctgtagggttctgctatggtactacacgcgcagggagagactgtctacc 329
DB 323 acctgaagggcctgtagggttctgctatggtactacacgcgcagggagagactgtctacc 382
QY 330 tctggaataatgtagcagcagcctgtagcctgtagcctgtagcctgtagcctgtagcct 389
DB 363 tctggaataatgtagcagcagcctgtagcctgtagcctgtagcctgtagcctgtagcct 420
QY 390 gataataacccactcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcct 449
DB 421 -----ctgagagctcattgtagcctgtagcctgtagcctgtagcctgtagcct 446
QY 450 tgatttaccatccctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcct 509
DB 447 tgatttaccatccctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcct 506
QY 510 cctgtaagaagaagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcct 569
DB 507 cctgtaagaagaagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcctgtagcct 566
QY 570 caagctcacacagcagagcgagcagcctcaacacgcgcacacgagcagcctgtagcctgtagcct 629
DB 567 tagcttcacacagcagagcgagcagcctcaacacgcgcacacgagcagcctgtagcctgtagcct 626
QY 630 ctgctcagacatagagggcctacacagcctgtagcctgtagcctgtagcctgtagcctgtagcct 689
DB 627 ctgctcagacatagagggcctacacagcctgtagcctgtagcctgtagcctgtagcctgtagcct 686
QY 690 tgcgtcagagcctcctcctcaatgctgctgtagcctgtagcctgtagcctgtagcctgtagcct 749
DB 687 tgcgtcagagcctcctcctcaatgctgctgtagcctgtagcctgtagcctgtagcctgtagcct 746
QY 750 cttccacaga 758
DB 747 cagcctgga 755

RESULT 4
AAH07034
ID AAH07034 standard; cDNA; 711 BP.
AC AAH07034;
DT 26-JUN-2001 (first entry)
DE Human cDNA clone (5'-primer) SEQ ID NO:3869.
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
PE 28-JUL-2000; 2000EP-0116126.
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
PA (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
DR
```

```
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 1; SEQ ID 3869; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
CC
SQ Sequence 711 BP; 149 A; 239 C; 209 G; 111 T; 3 other:

Query Match 38.1%; Score 418.8; DB 22; Length 711;
Best Local Similarity 78.7%; Pred. No. 4.7e-77;
Matches 581; Conservative 0; Mismatches 84; Indels 73; Gaps 4;

QY 7 ggcacagagcagcagcgtgtggtggtcagcagcgtgtagcgaacacagcagcgtggaag 66
DB 1 gacttgagcagcagcagcgtgtggtggtcagcagcgtgtagcgaacacagcagcgtggaag 60
QY 67 gagaagtcgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 126
DB 61 gaaagtcgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 120
QY 127 ctggcgcaactctgccccttgccgagcagcagcagcagcagcagcagcagcagcagcagc 186
DB 121 ctggcgcaactctgccccttgccgagcagcagcagcagcagcagcagcagcagcagcagc 180
QY 187 atgcgctcacacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 232
DB 181 atgcgctcacacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
QY 233 caggttggaagaaagggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 292
DB 241 caggttggaagaaagggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
QY 293 gtcagtgtactacacgcgcagagagagacatgcttactctgtagaagaaatgcaagcagc 352
DB 301 gtcagtgtgtacacgcgcagagagagacatgcttactctgtagaagaaatgcaagcagc 360
QY 353 ctggagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 412
DB 361 ctggagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 375
QY 413 ccaatttctcttgagagcattgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 472
DB 376 -----ctgagagcattgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 424
QY 473 aagaggaactcaagacgcctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 532
DB 425 aagaggaactcaagacgcctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 484
```



```

QY 533 ccccaacagagcgccactttctcctggaatgaagacagctacc-accagagagcg 591
    ||||| ||||| ||| ||| ||||| ||||| ||||| ||| |||||
Db 485 ccccaacagagcggtgtctctccttgcatgaagatgacccaagcgcgcgcg 544
QY 592 acgctcaactcaaaagcgccactggaaggtgtctgacgcctcgaagacataagggc 651
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 accctcaactcaaaagcgccactggaaggtgtctgacgcctcgaagacataagggc 604
QY 652 tacaagccccctgacagacttccctgcggagagccctgcgcag--cctccctgc 709
    ||||| ||||| ||| ||||| ||| ||||| ||| ||| |||
Db 605 tacaagccactgtgcaactcttcancatcggaaccctgactcaaancccccgcgca 664
QY 710 aatgcctggtgtctatct 727
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 atgcctggtgtctatct 682

RESULT 5
AAH03708 standard; cDNA; 756 BP.
XX
AC AAH03708;
XX
DE 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:543.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 543; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the

```

```

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 756 BP; 148 A; 263 C; 218 G; 123 T; 4 other:
Query Match 35.7%; Score 392.8; DB 22; Length 756;
Best Local Similarity 79.3%; Pred No. 1, ie-71;
Matches 571; Conservative 0; Mismatches 74; Indels 75; Gaps 6;
QY 44 ggtcgaacaccgagctgcyggaagagaagtcgcygagccgagccgagccgagcc 103
    ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 gttcgaaccagagctgcygcaagaagaatcccgagtgccgagccgagccgagcc 142
QY 104 aggaagacgaggtgtgtgtaccagctgcygcaactcgcctcttgccgcyggtacgg 163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 aggaagacgaggtgtgtgtaccagctgcygcaactcgcctcttgccgcyggtacgg 202
QY 164 cgaacctggaacagagcctcatatgcygctcacaactcagctacgtgcgcatgacgcgc 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203 ccaacctggaacagagcctcatatgcygctcacaactcagctacgtgcgcatgacgcgc 262
QY 224 tctgcgc-----agcaggttgaaagagggagagacactgtagcgcgtgt 269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 tctgcgcgcgagggagtggaaccaggttggaagagggagagaaacactgtagcgcgtgt 322
QY 270 acctgaagccctggaaggttctgtatgttaccacccgcggaggagaaatggtctaac 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 acctgaagccctggaaggttctgtatgttaccacccgcggaggagaaatggtctaac 382
QY 330 tgtcgaagaatgtcagaacagcctggtgcctcagatcagctctgtcctcctcct 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 tgtcgaagaatgtcagaacagcctggtgcctcagatcagctctgtcctcctcctcct 420
QY 390 gatacataaccacacctcctgttacaacttctcctggaagctcatgtgacaagatcct 449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 -----ctggaagctcatgtgacaagatcct 446
QY 450 tgatttatcatcctcctgtgaccagaaggaacttcaagaagccttgacccccaagccga 509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 tgatttatcatcctcctgtgaccagaaggaacttcaagaagccttgacccccaagccga 506
QY 510 cctgcaagaagaagcttgaagccccaacagagcgccacttccctgcaatgaaagag 569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 507 cctgcaagaagaagcttgaagccccaacagagcgccacttccctgcaatgaaagag 566
QY 570 cagctcaaccagagagcgacgctcaactcaactcaaaagcgagccactgtgaagtgtcga 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 567 taagctcaaccagcgcgcgccacccccaactcaaaagcgag-cacctggaaggtgtcga 625
QY 630 ctgtctaggaacatatgagggcgctacaagcc--cctgacaagacttccctgcggagggcc 688
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 ctgtctaggaacatatgagggcgctacaagccacacttccctgcggagggagggcc 685
QY 689 ctgcctcgagcctccc--ctgcaatgctgtgtgttatc-tgtgaagccatcccccagc 745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 ctgacccaacaaacccgnttgantgtcgtgtctatcttgtgaagcattcccaacc 745

RESULT 6
AAH15203 standard; cDNA; 1708 BP.
XX
AC AAH15203;
XX
DE 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13302.

```

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss  
Homo sapiens.  
EP1074617-A2.  
07-FEB-2001.  
28-JUL-2000; 2000EP-0116126.  
29-JUL-1999; 99JP-0248036.  
27-AUG-1999; 99JP-0300253.  
11-JAN-2000; 2000JP-0118776.  
02-MAY-2000; 2000JP-0183767.  
09-JUN-2000; 2000JP-0241899.  
(HELI-) HELIX RES INST.  
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.  
Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length CDNA defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length CDNA -  
Claim 8; SEQ ID 13302; 2537pp + CD ROM; English.  
The present invention describes primer sets for synthesizing 5602  
full-length CDNA defined in the specification, where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length CDNA. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length CDNA. The primers allow obtaining of the full-length  
CDNA easily without any specialized methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to  
AAB95693 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.  
Sequence 1708 BP; 340 A; 588 C; 461 G; 319 T; 0 other;

Query March	23.0%	Score 253.4	DB 22	Length 1708
Best Local Similarity	84.8%	Pred. No. 6.1e+3		
Matches 284	Conservative	0	Mismatches 51	Indels 0
			Gaps	0
QY	424	ctgagagcattgagacacagatattcttgattatcacccctgtgaccaaggaaact	483	
Db	281	ctgagagccattgagacacagcatcttctgatttcaaccacctgtgaccaaggagctt	340	
QY	484	caagagcgcctgagaccccccaagcgcgaacctgttcaaaagaagaactgtgaaagccccaacaga	543	
Db	341	caggaagcgcctgagaccccccaagcgcctgttcaggaggaaggtgtgagagcccccaagag	400	
QY	544	cgcacattttccctcgagatgaagagacgcctcaaccagagagaagggcagcgtcaactc	603	
Db	401	cgggtctctctctctgtgagatgaagatgcgtctcacacagcgcgaggcgacacctcaactc	460	

Oy		604	aaagcggccacctggaaggtgcctgcaactgctcaaggacatatgaggccaagaagccccct	663
Dd		461	aagcgccaccttggaagtgcttgcgaactgctcttcggacatatgagggccacaagaccact	520
Oy		664	gcaacagacttcccctgcggagagccctcgtctccgaagcctccctcgcaatgcttgtt	723
Dd		521	gcgcgagacttccagcttgtagagcccttcgaactagaagcccccgctgagtgtgtctc	580
Oy		724	atctgtgaagccatccccagactcccttcacaga	758
Dd		581	atctcgaaagccatccccaccaccagagagcttga	615
RESULT	7			
AH03186				
ID	AH03186	standard; cDNA; 873 BP.		
XX				
AC				
XX	AH03186;			
DT	26-JUN-2001	(first entry)		
DE	Human cDNA clone (5'-primer) SEQ ID NO:21.			
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss			
XX	Homo sapiens.			
OS	Homo sapiens.			
PN	EPI074617-A2.			
PD	07-FEB-2001.			
XX				
FE	28-JUL-2000; 2000EP-0116126.			
PR	29-JUL-1999; 99JP-0248036.			
PR	27-AUG-1999; 99JP-0300253.			
PR	11-JAN-2000; 2000JP-0118776.			
PR	02-MAY-2000; 2000JP-0183767.			
PR	09-JUN-2000; 2000JP-0241899.			
XX				
PA	(HELI-) HELIX RES INST.			
PI	Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;			
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;			
PT	WPI; 2001-318749/34.			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602			
PT	full-length cDNAs defined in the specification, and for the detection			
PT	and/or diagnosis of the abnormality of the proteins encoded by the			
PT	full-length cDNAs -			
XX				
PS	Claim 1; SEQ ID 21; 2537bp + CD ROM; English.			
CC				
CC	The present invention describes primer sets for synthesizing 5602			
CC	full-length cDNAs defined in the specification. Where a primer set			
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary			
CC	to the complementary strand of a polynucleotide which comprises one of			
CC	the 5602 nucleotide sequences defined in the specification, where the			
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination			
CC	of an oligonucleotide comprising a sequence complementary to the			
CC	complementary strand of a polynucleotide which comprises a 5'-end			
CC	sequence and an oligonucleotide comprising a sequence complementary to a			
CC	polynucleotide which comprises a 3'-end sequence, where the			
CC	oligonucleotide comprises at least 15 nucleotides and the combination of			
CC	the 5'-end sequence/3'-end sequence is selected from those defined in			
CC	the specification. The primer sets can be used in antisense therapy and			
CC	in gene therapy. The primers are useful for synthesising polynucleotides,			
CC	particularly full-length cDNAs. The primers are also useful for the			
CC	detection and/or diagnosis of the abnormality of the proteins encoded by			
CC	the full-length cDNAs. The primers allow obtaining of the full-length			
CC	cDNAs easily without any specialised methods. AAH03186 to AAH13628 and			
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to			
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632			

CC represent oligonucleotides, all of which are used in the exemplification  
of the present invention.

XX Sequence 873 BP; 175 A; 305 C; 248 G; 142 T; 3 other;

Query Match 23.0%; Score 253.2; DB 22; Length 873;  
Best Local Similarity 86.6%; Pred. No. 5.7e-43;  
Matches 279; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 424 ctgagctcatggaacagatcttattatccatccctgtgaccaagagaactt 483  
DB 281 ctgagctcatggaacagatcttattatccatccctgtgaccaagagaactt 340  
QY 484 caagagcgcctgaaccccaagcgcctgtcaagaagaagctggaagcccaagag 543  
DB 341 cagagcgcctgaaccccaagcgcctgtcaagaagaagctggaagcccaagag 400  
QY 544 cgcacatttccctcgaatgaagagcagcctcaccaagaagagcgagcctcaaccc 603  
DB 401 cgtgtcttctctcgtgagctgaagagtaagctcaccaagcgcgcgcacccctcaaccc 460  
QY 604 aagagcgccacctggaaggtgtcgtcactgctcaagacatatgaggcctacaagccct 663  
DB 461 aagagcgccacctggaaggtgtcgtcactgctcgtgacatatgaggcctacaagccct 520  
QY 664 gacagagcttccctcgc 723  
DB 521 ggcgcagacttctcagctctggaagcctgtgacagagcccgctgagtgctgtgtc 580  
QY 724 atctgtgaagccatcccccagc 745  
DB 581 atctgtgaagccatcccccagc 602

## RESULT 8

AAH03810  
ID AAH03810 standard; cDNA; 733 BP.

XX  
AC AAH03810;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:645.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length cDNAs -

PS Claim 1; SEQ ID 645; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination  
of the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.

XX Sequence 733 BP; 143 A; 264 C; 204 G; 119 T; 3 other;

Query Match 19.2%; Score 211.6; DB 22; Length 733;  
Best Local Similarity 83.1%; Pred. No. 2e-34;  
Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 472 caagaggaacttcaagagcgccttgaccccaagcgcgaacttcaagaagaagctgaa 531  
DB 272 caagaggaacttcaagagcgccttgaccccaagcgcgaacttccagaggaagtgag 331  
QY 532 gccccaagagcgcgcacttccctcgcgaatgaagaagcgcctcagcagagggcgc 591  
DB 332 gccccaagagcgcgcacttccctcgcgaatgaagaagcgcctcagcagagggcgc 391  
QY 592 acgctcaacctcaagaagcgcgcacttgaagaagtgctgactgctcagacatatgagggc 651  
DB 392 accctcaacctcaagaagcgcgcacttgaagaagtgctgactgctcagacatatgagggc 451  
QY 652 tacaagcccccctgcaagaagcttccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 711  
DB 452 tacaagcccccctgcaagaagcttccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 511  
QY 712 tgctgtgtctatctgtgaagccatcccccagctcccttcaagatg 761  
DB 512 tgctgtgtctatctgtgaagccatcccccagctcccccagcagcagcttgaggg 561

## RESULT 9

AAH18364  
ID AAH18364 standard; cDNA; 1841 BP.

XX  
AC AAH18364;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:18402.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishi S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8: SEQ ID 18402; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX  
SQ Sequence 1841 BP; 404 A; 514 C; 519 G; 404 T; 0 other:  
  
Query Match 19.2%; Score 211.6; DB 22; Length 1841;  
Best Local Similarity 83.1%; Pred. NO. 2.4e-34;  
Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
  
OY 472 caaaggaactcaagaagccctgaccccgagccgaacctgtcaagaagaagctggaa 531  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 272 caaagagagactcagagccctgaccccgagcccgagccctgtcaggaagaagtgag 331  
  
OY 532 gccccaacagagcgccacttccctcgcaatgaagaagacagctcaccagcagagcgcc 591  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 332 gccccaacagagcggtgtcttctcctgagcagtaagaagtaagcaccagcgcgagcgcc 391  
  
OY 592 accgtcaacctcaaaagcgccacgtgagaggtgtcgtcagctcgaagacatagagggcc 651  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 392 accctcaacctcaaaagcgccacgtgagaggtgtcgtcagctcgtgacatagagggcc 451  
  
OY 652 tacaagcccttgacagacttccctcgtcgagagccctcgtccgaagcctccctcgaa 711  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 452 tacaagcaccctgcagacttctccagctcgtgagagccctgagcagcccggtgag 511  
  
OY 712 tgctgtgtcttctgtgaagcattcccccctccctccacagatgg 761  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 512 tgctgtgtcttctgtgaagcattcccccctccctccacagatgg 561  
  
RESULT 10  
AAK58981

ID AAK58981 standard; cDNA; 2730 BP.  
XX  
XX AAK58981;  
AC  
XX 23-AUG-1999 (first entry)  
DT  
XX  
XX Human transcription regulator MOP2 cDNA.  
DE  
XX MOP2; member of the PAS superfamily; bHLH-PAS; human;  
KM transcription regulator; hypoxia inducible factor 2 alpha; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 33..2645  
FT CDS /\*tag- a  
FT  
XX  
XX MN9928464-A2.  
PN  
XX  
XX 10-JUN-1999.  
PD  
XX  
XX 27-NOV-1998; 98WO-US25314.  
PF  
XX  
XX 28-NOV-1997; 97US-0066863.  
PR  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA  
XX  
XX Bradford CA, Gu YZ, Hogenesch JB;  
PI  
XX  
XX WPI; 1999-371120/31.  
DR  
XX  
XX P-PSDB; AAY06290.  
DR  
XX  
XX Developmental signal transduction associated proteins  
PT  
XX  
XX  
XX  
XX Claim 6; Page 94-95; 106pp; English.  
PS  
XX  
XX This is the nucleotide sequence of a cDNA encoding MOP2 (see  
CC AAY06290), a novel member of the PAS superfamily, where PAS strands  
CC for PBR/ARNT/SIM domains. The cDNA was identified in an iterative  
CC search of human ESTs designed to identify basic-helix-loop-helix-PAS  
CC (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)  
CC or the Ah receptor nuclear translocator (ARNT). To obtain extended  
CC open reading frames for each EST, an anchored-PCR strategy was used  
CC to amplify additional flanking sequences from a commercial HepG2  
CC library. MOP2 appears to be related structurally and functionally  
CC to MOP1 (see AAY06289). It interacts with ARNT, but not AHR, and  
CC drives transcription in its ARNT-dimerised form. Unlike MOP1, it  
CC does not appear to interact significantly with HSP90. MOP2 is  
CC induced by low oxygen and may be involved in hypoxia responses in  
CC difference cells and tissues. MOP2 is sometimes referred to as  
CC hypoxia inducible factor 2 alpha. The invention provides novel  
CC MOPs 2-9 nucleic acids (see AAK58981-88) and proteins (see AAY06289-97).  
CC These are useful in a variety of research, diagnostic and  
CC therapeutic applications. Several of the MOPs are alpha-class  
CC hypoxia-inducible factors. Others are involved in circadian signal  
CC transduction.  
CC  
XX  
SQ Sequence 2730 BP; 659 A; 839 C; 727 G; 505 T; 0 other:  
  
Query Match 15.9%; Score 174.6; DB 20; Length 2730;  
Best Local Similarity 57.5%; Pred. NO. 1.1e-26;  
Matches 446; Conservative 0; Mismatches 244; Indels 85; Gaps 4;  
  
OY 2 aatcgacagagcgcatgctgttggtgctgcagcgctgaggtcgaacaccagactgc 61  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 16 aaagggcaccagcgcaaatgacatgacagctgacaaaggaagaagtagctcggagagga 75  
  
OY 62 ggaaggaagtcgaggagcgcgcccgagcgccgagcagcaggaagagagtgctgt 121  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 76 ggaaggaagtcgaggagcgcgcccgagcgccgagcaggaagagagtagtctct 135  
  
OY 122 accagctgagcagactctgccccttgctgcgagcgctgacgagcagacttgacaaggcct 181

Db	136	atgagctcgcccaatgagctctgctctccccaacagtgtgagctccocatctgtaacaagcct	195
OY	182	ccatcatatgagcttcaacaactcagctctgcgacatgacacgcgctctctgcgcacag	235
Db	196	ccatcatatgagctatgagctcagctctctctgtaacaacaacaagcctcctctctcagttgct	255
OY	236	-----gtgaaaaaggggagagccacacggagcgctctgtactatgaagccc	281
Db	256	ctgaaaacgagctccgaagccggaagctcacagcagatgtgacaaactgtacactgaaagcct	315
OY	282	tggagagttctgtcatatgtactcaacgcgagaggagacatgcttacccttgcgaaatg	341
Db	316	tggagaggtttcatatgtgcctgtgtgagccaaagatgagcagatcttctctgtacagaanaa	375
OY	342	tcagacagacactctggcgctcaagtcagtcgagacctctgtctctccctctgtataaaccc	401
Db	376	tcagagagttcatatgtgagcttaacag-----	401
OY	402	cactctgtgtacaaattctctctgtgagcctcatatgtgacacagtatcttgaattatcca	461
Db	402	-----gtgagagctaaacagacataagatacttgaattactta	439
OY	462	tcctctgtgaccaaagaggaacttcaagacgcttgaacccca-----ggcgagacctgtc	515
Db	440	tcctctgtgacaaatgagagagatctgtgtgaaacactgagcttcaaaaatgtgcctctgttgg	499
OY	516	aaagaaagaaagctgtgaaagccccaacaagaagcgcaacttctccctcgaaatgaagaagcgt	575
Db	500	gaaaaaaagacaagaacatgttccacaagcgaggaacttctcatgagatgagatggaagcgt	559
OY	576	cacacagcagagagggcgacgctccaaacccccaagaagcggaactgtgaaggtgtctgacgtgc	635
Db	560	caccaacacgaagcgctacatctgtccaacccccaagtcagccaactgtgaaggtctctgacgtcac	619
OY	636	agagacatatgtgggctctacaaagcccccctgtacagagacttcccctggccggagagcctctgc	695
Db	620	gggcacaggttgaagttctacaacaacatcgccctc---ctcacataagtcgtgtgtctacaa	676
OY	696	cgagccttcctctgaaagcctctgtgtctatctgtgaagacatcccacagctccc	750
Db	677	ggaaagccctctgtctctgtcccatcatcatatgtgtgaaccatcatcagacattcc	731

```

RESULT 11
AAZ94052
ID AAZ94052 standard; DNA; 1607 BP.

```

DT 05-JUN-2000 (first entry)

Human endothelial PAS domain protein-1 dominant negative mutant DNA.

KW Endothelial PAS domain protein-1; EPAS1; angiogenesis;

gene therapy; human; dominant negative mutant; ss.

OS Homo sapiens.

XX XX

FT	150..1610
CDS	

[illegible]

2  
1  
1  
1  
0  
0  
0  
0

XX  
2

**XX**

XX

PA (HARD ) HARVARD COLLEGE.

PI Lee M, Maemura K, Hiesh C;

DR WPI; 2000-205996/18.

**XX**

PT atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue

PT circulation

PS Disclosure; Page 4-5; 57pp; English.

The present sequence is that of nDNA encoding a dominant negative mutant (see AAY79160) of human endothelial PAS domain protein-1 (EPAS1). The mutant lacks the transactivating domain (see AAY79158) of the wild-type protein (see also AAY79161). The invention is based on the discovery that EPAS1 binds to cis-acting regulatory sequences associated with genes encoding angiogenic factors such as vascular endothelial cell growth factor (VEGF) and VEGF receptors such as KDR/Flk-1 and flt-1, thereby transactivating the promoters of such genes. A claimed method of inhibiting angiogenesis in a mammal comprises administering to the mammal a compound which inhibits binding of EPAS1 to the cis-acting transcription regulatory DNA of an angiogenic factor (see AAY24051). The compound may be an EPAS1 polypeptide lacking a transactivating domain or a nucleic acid encoding such a polypeptide. When such an EPAS1 mutant is bound to a cis-acting regulatory DNA, it prevents wild-type EPAS1 binding and thereby inhibits transcription of a gene encoding an angiogenic factor such as VEGF, and hence angiogenesis. The mutant is preferably administered to an atherosclerotic lesion or to a tumour site. Angiogenesis is also inhibited using a compound, such as an EPAS1 dominant negative mutant, which inhibits binding of EPAS1 to the EPAS-binding element, ARNT4 (see AAY79162).

Sequence 1607 BP; 395 A; 480 C; 442 G; 290 T; 0 other;

Query Match	15.78;	Score 173;	DB 21;	Length 1607;
Best Local Similarity	57.48;	Pred. No. 2.1e-26;		
Matches 445; Conservative	0;	Mismatches 245;	Indels 85;	Gaps 4;

[illegible]



XX AA294053;  
AC  
XX  
XX 05-JUN-2000 (first entry)  
DE DNA encoding human endothelial PAS domain protein-1.  
XX  
XX Endothelial PAS domain protein-1; EPAS1; human; angiogenesis;  
KM antiatherosclerotic; antitumour; atherosclerosis; tumour;  
KM vascular disease; vulnerrary; cardiant; vasotropic;  
XX cerebroprotective; gene therapy; ss.  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 150..2762  
ET /tag= a  
XX  
XX WO200009657-A2.  
XX  
XX 24-FEB-2000.  
XX  
XX 13-AUG-1999: 99WO-US18539.  
XX  
XX 14-AUG-1998: 98US-0096515.  
XX  
XX (HARD ) HARVARD COLLEGE.  
XX  
XX Lee M, Maemura K, Hlesh C;  
XX  
XX WPI: 2000-205996/18.  
XX P-PsDB: AAY79161.  
XX  
XX  
XX Modulation of angiogenesis in mammals, useful for treating e.g.  
XX atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue  
XX damage, ischemia, balloon angioplasty, frostbite, gangrene or poor  
XX circulation  
XX  
XX  
XX Disclosure: Page 14-15: 57pp: English.  
XX  
XX  
XX The present sequence is that of cDNA encoding human endothelial PAS  
XX domain protein-1 (EPAS1, see AAY79161). The invention is based on  
XX the discovery that EPAS1 binds to cis-acting regulatory sequences  
XX associated with genes encoding angiogenic factors such as vascular  
XX endothelial cell growth factor (VEGF) and VEGF receptors such as  
XX KDR/Flk-1 and Flt-1, thereby transactivating the promoters of such  
XX genes. A claimed method of inhibiting angiogenesis in a mammal  
XX comprises administering to the mammal a compound which inhibits  
XX binding of EPAS1 to the cis-acting transcription regulatory DNA of  
XX an angiogenic factor (see AA294051). The compound may be an EPAS1  
XX polypeptide lacking a transactivation domain (see AAY79160) or a  
XX nucleic acid encoding such a polypeptide, an antisense nucleic acid  
XX complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody.  
XX The compound is preferably administered to an atherosclerotic  
XX lesion or to a tumour site. Angiogenesis is also inhibited using a  
XX compound, such as an EPAS1 dominant negative mutant, which inhibits  
XX binding of EPAS1 to the EPAS-binding element, ARRY4 (see AAY79162).  
XX Angiogenesis can be promoted by administering EPAS1 DNA to increase  
XX expression of VEGF or VEGF receptor in endothelial cells of a  
XX patient suffering from peripheral vascular disease, cerebral  
XX vascular disease, hypoxic tissue disease (e.g. hypoxic damage to  
XX heart tissue), or coronary vascular disease as well as to treat  
XX patients who have, or have had, transient ischemic attacks,  
XX vascular graft surgery, balloon angioplasty, frostbite, gangrene,  
XX or poor circulation.  
XX  
XX  
XX Sequence 2818 BP; 669 A; 882 C; 764 G; 503 T; 0 other;

Query Match 15.7%; Score 173; DB 21; Length 2818;  
Best Local Similarity 57.4%; Pred. No. 2.3e-26;  
Matches 445; Conservative 0; Mismatches 245; Indels 85; Gaps 4;

OY 2 aattcgacacgagggccatgctgttggtcagcgctgagctgcaaccagctgc 61  
DB 133 aaagggccacacgcgcacatgacagctgacaaagagaagaagagtagctggagga 192  
OY 62 ggaaggaagaagctcgaggagcgcccgacgcgagcagcagagagagagtgctgt 121  
DB 193 ggaaggaagaagctccggtatgctgcggtgctgcgagcagagagagagtgctgt 252  
OY 122 accagctgagcacaactctgctccttgcgcggtgctgacgcgcacctgacaagcct 181  
DB 253 atgagctggcccatgagctgctcctcccaagtgtagctcccatctgacaagcct 312  
OY 182 ccatactgctgcacacatgactgctgcatgcaacgcctctgagcagcag----- 235  
DB 313 ccatactgctgcagacatgactgctgcatgcaacacacacacagctcctcctcagttgct 372  
OY 236 -----gtgaaaaaggggagagacactgtagcgcctgtacactgaaagcct 281  
DB 373 ctgaaaacgagctcgaaagcgcgaagctgacacgacagatggaacactgttactgaaagcct 432  
OY 282 tggaggggttcgtcatgctacacgcgagaggaagacatgcttaccgtgcgaaagt 341  
DB 433 tggaggggttcgtcatgctacacgcgagaggaagacatgcttaccgtgcgaaagt 492  
OY 342 tcagcaagcagcctggggtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 401  
DB 493 tcagcaagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 518  
OY 402 cactcctgtagcaccatctcctcgtgagctcatgtagacagatatttattatcca 461  
DB 519 -----gtgagctaaagcgagacatgacttcttgacttactca 556  
OY 462 tccctgtgaccagaaggaacttcaagaagcctgacccca-----ggccgaactgtc 515  
DB 557 tccctgtgaccagaaggaacttctgtagaacctgagcttcaaaaatgctgtgtttg 616  
OY 516 aaagaagaagctggaagcccaagaagcgccacttctcctcgtgaatggaagcagct 575  
DB 617 gaaaaaaagcaagaagcagctcccaagaagcgagcagcttctcagtagagtagcagagct 676  
OY 576 caccagaagaagcgagcagctcaccacccaagcgccacactggaagtgtagctagctc 635  
DB 677 caccagaagaagcgagcagctcaccacccaagcgagcagctggaagtgtagctagctc 736  
OY 636 aggaacataagagggcctacaagccctgcagcagactcctcgtcgagagcctgctc 695  
DB 737 gggccaggtgaaggtctacacacactgcccctc---ctcaacatagctgtgtgtctcaa 793  
OY 696 cggagcctccctcgaatgctgt 750  
DB 794 ggaagcctgctgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 848

RESULT 14  
ID ABA14358 standard; DNA: 32169 BP.  
AC ABA14358;  
XX  
XX  
XX 23-JAN-2002 (first entry)  
DE Human nervous system related polynucleotide SEQ ID NO 6689.  
XX  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;  
XX immunosuppressive; antihistaminic; anti-HIV; antibacterial; vulnerrary;  
XX antiparkinsonian; antisclerotic; antianemic; antiatherosclerotic; cancer;  
XX antineoplastic; hepatoprotective; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
XX Homo sapiens.  
XX



PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-022557.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251997.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA - ) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;



Oy 503 -----ggccgaacctgtcaagaagaagcttgaagccccacaagagcgccactttccc 556  
Db 491 aaatgacctgtgttttggaaaaaagcaaatgttccacagagcggtacttctca 550  
Oy 557 tgcgatatgaagacgcttccaccagagagggcgacgctcaacctcaagcgccacct 616  
Db 551 tgaagatgaagtgcaaggtcaccacaagagccgtactgttcaacctcaagtcagccacct 610  
Oy 617 ggaaggtctgcactgtcaagacatatgaaggcttacaagccccctgtcacagacttccc 676  
Db 611 ggaaggtcttgcactgtcacagggcgagtgaaagtctacaacaactgcctc---ctcaaca 667  
Oy 677 ctgccggagacctcgtcccgagcctccccctgcaatgacctgtgtcttatctgtgaagcca 736  
Db 668 atagctgtgtgtacacaagagccccctgtctcctgtcctcatcatatgtgtgaaccaa 727  
Oy 737 tccccagctcccc 750  
Db 728 tcagcaccacalc 741

Search completed: September 8, 2002, 06:49:31  
Job time: 6984 sec

**Dy**    2    aatcgcgacagaggccatcgctgtggcgctgcagtcgctgaagtgcgaaccgcagctgc 61  
         || | ||| | |||| | || || | || || || |  
**Db**    133 AAGGCGCACACGCCACAATGCACGCTGCACAGAGAGAAGAAAGAGTTCGGAGAGGA 192

164968/20

[illegible]

RESULT 2  
 US-08-785-241-2  
 Sequence 2, Application US/08785241  
 Patent No. 56595963  
 GENERAL INFORMATION:  
 APPLICANT: McKnight, Steven L.  
 APPLICANT: Russell, David W.  
 APPLICANT: Tian, Hui  
 TITLE OF INVENTION: Endothelial PAS Domain Protein  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SCIENCE & TECHNOLOGY IAW GROUP  
 STREET: 268 BUSH STREET, SUITE 3200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

```

1 APPLICATION NUMBER: US/08/785,241
2 FILING DATE: 17-JAN-1997
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: OSMAN, RICHARD A
6 REGISTRATION NUMBER: 36,627
7 REFERENCE/DOCKET NUMBER: UTSD:1222
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: (415) 343-4341
10 TELEFAX: (415) 343-4342
11 INFORMATION FOR SEQ ID NO: 2:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 3031 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: double
16 TOPOLOGY: linear
17 MOLECULE TYPE: CDNA
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

Query Match	14.28;	Score 156.6;	DB 1;	Length 3031;
Best Local Similarity	57.18;	Pred. No. 6e-28;		
Matches 441; Conservative	0;	Mismatches 244;	Indels 88;	Gaps 5

[illegible]



[illegible]

```

RESULT      5
US-09-148-547-1
: Sequence 1, Application US/09148547
: Patent No. 6124131
: GENERAL INFORMATION:
: APPLICANT: Semenza, Gregg L.
: TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use
: FILE REFERENCE: 07265/151001
: CURRENT APPLICATION NUMBER: US/09/148,547
: CURRENT FILING DATE: 1998-08-25
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3736
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (23)..(2509)
: US-09-148-547-1

```

Query Match	12.2%;	Score 134.6;	DB 3;	Length 3736;
Best Local Similarity	54.7%;	Pred. No. 1e-22;		

[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/235,217





Accession	Sequence	Position
Oy	ccatccccgtgtaaccaagaagacttaagatgcctctgaacccccagcgccgaactgltcaaa	518
Db	tcattcatgtatccatcgtagaggaatagagaatctttacacacaggaatgacctgttgaa	495
Oy	gaagaagcttgaagcccccacagaacgcacactttccctcgaatgaagacacgtcac	578
Db	aaaggggttaaaacaaaaacacacacgcgaactttttctcagatgaagtgtaacccaac	555
Oy	cagcgaagagcgacgcgtcacaaccacaagaacgagccacttgaaagtgctgacgtcag	638
Db	taaccgaggaagaaactatgaacatgaatctgcacaacatggaaggtatgtgacacgcacagg	615
Oy	acatatgaagggcctcaacaagccccctcgcacagacttcccctgcgcggaagccctgcgtcga	698
Db	ccacattcaccgtatatgatatacca-----acagtaacccaacctcagtggtgataaaga	669
Oy	ggcttcacctgaatgctcgtggtgtctatctgtgaagccatccccagc	745
Db	accacattatgacctgcttggtggctgattgttgaaacccatttcctaac	716

```

RESULT 8
US-08-785-310A-1
Sequence 1, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-Jan-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSJ:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2078 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-785-310A-1

```

Query Match	8.1%;	Score 89.6;	DB 2;	Length 2078;
Best Local Similarity	67.9%;	Pred. No. 3,3e-12;		
Matches 125;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;

[illegible]

Db 370 GAGTTCCTTCGAGCTGGGCGCAACCTCTCTCCGCTGGCCGGCGGCATCTCCATTCAGCTGAC 429

Oy 175 aaggcctccatcatalgagcgcctcaacaataagctactctgcatgacgcgcctctgagca 234

Db 430 AAGCTTCATCATGTCGGCCTCAGCTACCATCTCCGCGCCGCCGTTCCGCGCCTG 489

Oy 235 ggtg 238

Db 490 GGGG 493

```

US-08-785-310A-2
9
Sequence 2, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
TITLE OF INVENTION: Neutrolin PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-Jan-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSID:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA
IS-08-785-310A-2

```

Query Match	6.58;	Score 72;	DB 2;	Length 2082;
Best Local Similarity	62.08;	Pred. No. 4.7e-08;		
Matches 114;	Conservative 0;	Mismatches 70;	Indels 0;	Gaps 0;

QY	55	gagcttgaggaggaagtctgcggagacgagcccgacgagcgagcgacgaggaagagag	114
	190	GGCGAGGGCAAAAGAAAGTCCCGCAACGGTCGCGCGCTGGCGAGCGGGGAAGAAATCTG	249
Db	115	gtgtgtgtaacagactgagcaactctgccccttttggcgcggtctaaagcgacactggagac	174
QY	250	GAGTTCCTTCAGGCTAGGCCAAGCTGCTCCTCTGTGCGGAGTCCATTCTCAAGCCAGCTGGAC	309
Db	175	aagagctcatcatatgcgctcacaatcagcttacctgacgatgacagcgctctgcagaga	234
QY	310	AAGGCGTCACATCGTGGCTCTTAAGGTCACCTACTCTCGCCTCGGTGGTTTGGCGGCGTG	369
Db	235	ggtg	238
QY	370	GGGG	373
Db			

```

      TITLE OF INVENTION: Process for Its Production
    NUMBER OF SEQUENCES:   12
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPILER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/ms-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,349
FILING DATE: 25-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/364,439
FILING DATE:
APPLICATION NUMBER: US 08/112,061
FILING DATE: 26-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barker, M. P.
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 02481.1323-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Mus musculus
STRAIN: osteoblastic cell line MC3T3E1
FEATURE:
NAME/KEY: CDS
LOCATION: 284..2671
US-08-738-349-1

Query Match.           5.7%; Score 62.8; DB 2: length 3581;
Best Local Similarity 81.1%; Pred. No.8.6e+06;
Matches 73; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY  1011 aaagacatcttcctgttaggcaccagaataaaaaaaaaaaaaaaaaaaaaaa
          ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB  3486 AAAAAAAGATCTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                                |||||               3545

OY  1071 aaaaaaaaaaaaacaactgcgccgc 1100
          |||||||||              |||||             3545
DB  3546 AAAAAAAAAAAAAAAAAAACGCGCCGC 3575


RESULT 12
US-08-722-126A-4
Sequence 4, Application US/08722126A
Patent No. 6034227
GENERAL INFORMATION:
APPLICANT: PECHT, Israel
APPLICANT: GUTHMANN, Marcelo D.
APPLICANT: TAL, Michael
TITLE OF INVENTION: A DNA MOLECULE ENCODING A MASt CELL
TITLE OF INVENTION: FUNCTION-ASSOCIATED ANTIGEN (MAFR)
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
```

```

: STREET: 419 Seventh Street N.W., Ste. 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 20004
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/722,126A
: FILING DATE: 08-OCT-1996
: CLASSIFICATION: 536
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04258
: FILING DATE: 06-APR-1995
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: IL 109257
: FILING DATE: 08-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: PECHT-1A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 737-3528
: TELEFAX: (202) 628-5197
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1461 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 54..617
: US-08-722-126A-4
:
: Query Match 5.7%; Score 62.2; DB 3; Length 1461;
: Best Local Similarity 71.3%; Pred. No. 8.4e-06;
: Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: QY 977 ccagaaaccacaatgtctcaaaaccacataaagacccctctgttgagccag 1036
: DB 1343 CCTTAAATTTAAAGAGTCCCAATTTAAATTAATTCCTTTTCGCTAAAAA 1402
:
: QY 1037 agaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 1091
: DB 1403 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1457
:
: RESULT 13
: PCT-US95-04258-4
: Sequence 4, Application PC/TUS9504258
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: A DNA MOLECULE ENCODING A MAST CELL
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NETMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04258
: FILING DATE: 06-APR-1995
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: IL 109257
: FILING DATE: 08-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: PECHT-1 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1461 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 54..617
: PCT-US95-04258-4
:
: Query Match 5.7%; Score 62.2; DB 5; Length 1461;
: Best Local Similarity 71.3%; Pred. No. 8.4e-06;
: Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: QY 977 ccagaaaccacaatgtctcaaaaccacataaagacccctctgttgagccag 1036
: DB 1343 CCTTAAATTTAAAGAGTCCCAATTTAAATTAATTCCTTTTCGCTAAAAA 1402
:
: QY 1037 agaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa 1091
: DB 1403 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1457
:
: RESULT 14
: US-09-227-357-71
: Sequence 71, Application US/09227357
: Patent No. 6342581
: GENERAL INFORMATION:
: APPLICANT: Fischer et al.
: TITLE OF INVENTION: 123 Human Secreted Proteins
: FILE REFERENCE: P2010P1
: CURRENT APPLICATION NUMBER: US/09/227,357
: CURRENT FILING DATE: 1999-01-08
: EARLIER APPLICATION NUMBER: PCT/US98/13684
: EARLIER FILING DATE: 1998-07-07
: EARLIER APPLICATION NUMBER: 60/051,926
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,793
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,925
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,929
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,803
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/052,732
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,931
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,932
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,916
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,930
: EARLIER FILING DATE: 1997-07-08
: EARLIER APPLICATION NUMBER: 60/051,918

```

EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 71  
LENGTH: 413  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (343)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (385)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (410)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-227-357-71

Query Match 5.6%; Score 61.8; DB 4; Length 413;  
Best Local Similarity 79.1%; Pred. No. 6.4e-06;  
Matches 72; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1001 aaaccacataagacgtctctgttaggcacagagaaaaaa 1060  
DB 315 aaatacatatcactcgtctgttanaaaaaaa 374  
QY 1061 aaaaaaaaaaaaaaaaaaaaaaaaaa 1091

DB 375 aaaaaaaaaaaaaaaaaaaaaaaaaa 405

RESULT 15  
US-08-545-196B-10  
Sequence 10, Application US/08545196B  
Patent No. 6080577  
GENERAL INFORMATION:  
APPLICANT: MUMNICH, JUDITH  
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,196B  
FILING DATE: 19-OCT-1995  
CLASSIFICATION: A35  
ATTORNEY/AGENT INFORMATION:  
NAME: FARACI, C. J.  
REGISTRATION NUMBER: 32,350  
REFERENCE/DOCKET NUMBER: 2121-110P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1582 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-545-196B-10

Query Match 5.5%; Score 60.4; DB 3; Length 1582;  
Best Local Similarity 74.5%; Pred. No. 2.3e-05;  
Matches 76; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 990 aaatgtctcacaaccacataagacgtctctgttaggcacagagaaaaaa 1049  
DB 1449 ATATGTTTAAAGTATATATATATATATTTTAAAAA 1508  
QY 1050 aaaaaaaaaaaaaaaaaaaaaaaaaa 1091  
DB 1509 AAAAAAAAAAAAAAAAAAAAAAAAAA 1550

Search completed: September 8, 2002, 05:58:37  
Job time: 6385 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 04:03:27 : Search time 1715.42 Seconds  
(without alignments)  
8654.820 Million cell updates/sec

Title: US-09-896-791b-2  
Perfect score: 1100  
Sequence: 1 gaattcgagcagcgccat.....aaaaaaaaacatcgcgccgc 1100

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	507.6	46.1	687	10	BF722881 mab24h10.
2	461	41.9	461	9	AL322407 m119b08.y
3	449.6	40.9	905	9	AL528423 AL528423
4	444.8	40.4	856	9	AL519496 AL519496
5	427	38.8	919	9	AL541807 AL541807
6	408.2	37.1	767	10	BG696633 BG696633
7	392.8	35.7	756	9	AU118510 AU118510
8	386.6	35.1	460	9	AA028416 AA028416
9	336	30.5	633	10	AA041707 m119b08.y
10	299.6	27.2	459	9	AA041707 m119b08.y
11	298.2	27.1	418	10	BM118906 BM118906
12	298.2	27.1	425	10	BM118909 BM118909
13	298.2	27.1	610	10	BM119659 BM119659
14	276.8	25.2	396	10	BF653212 BF653212
15	253.2	23.0	873	9	AU116888 AU116888
16	250.2	22.7	620	10	BF65142 OVA-NM114
17	244	22.2	865	9	AL535689 AL535689

c	18	242.8	22.1	417	9	AA988550	AA988550	UI-R-CO-1
c	19	242.8	22.1	486	9	AL137683	AL137683	UI-R-CO-h
c	20	239.8	21.8	421	9	AM520329	AM520329	UI-R-BUOP
c	21	239	21.7	321	9	AL171316	AL171316	UI-R-AFI-
c	22	233.2	21.2	279	9	AV322372	AV322372	AV322372
c	23	226.2	20.6	384	9	AV578946	AV578946	UI-R-AAO-
c	24	211.6	19.2	733	9	AU118844	AU118844	AU118844
c	25	196.8	17.9	353	9	AT510204	AT510204	m102907.y
c	26	191	17.4	430	10	BF654293	BF654293	278403 MA
c	27	188.6	17.1	228	10	BM483179	BM483179	536412 MA
c	28	174.6	15.9	858	9	AL544018	AL544018	AL544018
c	29	153.4	13.9	469	10	BF548129	BF548129	UI-R-AO-b
c	30	150.2	13.7	632	9	AM133666	AM133666	f110c03.y
c	31	141.2	12.8	662	10	BM182143	BM182143	f54b05.y
c	32	139.8	12.7	1657	11	AK017853	AK017853	Mus muscu
c	33	136.6	12.4	549	10	BE721988	BE721988	189855 MA
c	34	136	12.4	659	10	BM182886	BM182886	f62h07.y
c	35	136	12.4	905	9	AL549361	AL549361	AL549361
c	36	133.6	12.1	878	9	AU133294	AU133294	AU133294
c	37	132.4	12.0	769	10	BG72697	BG72697	602720844
c	38	128.4	11.7	430	9	AM211226	AM211226	u129b12.y
c	39	124.8	11.3	678	9	AL642308	AL642308	AL642308
c	40	123.6	11.2	533	9	AL877949	AL877949	fcs5e08.y
c	41	123.6	11.2	642	10	BM157049	BM157049	f43g02.y
c	42	123.6	11.2	758	10	BI890667	BI890667	z6537-2-0
c	43	120.2	10.9	526	10	BG884123	BG884123	f672b10.y
c	44	120.2	10.9	959	9	AL530209	AL530209	AL530209
c	45	119.6	10.9	833	9	AL552246	AL552246	AL552246

#### ALIGNMENTS

RESULT 1  
LOCUS BF722881 687 bp mRNA linear EST 03-JAN-2001  
DEFINITION mab24h10.y1 Soares.NMEBA branchial arch Mus musculus cDNA clone  
IMAGE:3971562.5' similar to TR:Q9QX54 Q9QX54 HYPOXIA-INDUCIBLE  
FACTOR 3 ALPHA ;, mRNA sequence.

ACCESSION BF722881  
VERSION BF722881.1 GI:12023883  
KEYWORDS  
SOURCE  
ORANISM  
house mouse.  
Mus musculus.

REFERENCE  
AUTHORS  
TITLE  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index.

JOURNAL  
COMMENT  
Unpublished (1997)  
Other ESTs: mab24h10.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-remail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonalço

Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
image.lnl.gov/image/html/lresources.shtml

MGI:1471594  
Seq primer: -40RP from Glibco  
High quality sequence stop: 469.  
Location/Qualifiers

FEATURES  
SOURCE  
1..687  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3971562"  
/clone\_id="Soares.NMEBA\_branchial\_arch"  
/tissue\_type="branchial arches"  
/dev\_stage="embryo, 10.5 dpc"



```

//log host="DH10B (phage resistant)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTATCCCAATCTGAAGTGGGAGCGCGCATGCAATTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Patima
Bonafido."
BASE COUNT      149 a      195 g      111 t      1 others
ORIGIN

```

Query Match	46.1%	Score 507.6	DB 10	length 687
Best Local Similarity	88.3%	Pred. No. 6	8e-54	
Matches 603, Conservative	0	Mismatches	10	Indels 70; Gaps 2;

**Dy** 82 gcgagcccgacgcccgcgcagcaggagaagagtgtctgtaccagctggcgcacactctg 141  
**Db** 1 GCGGCCGCGACCCGGCGCAGCACGAGACGGAGTGCTTACACACTGGCCACACACTTG 60

[illegible]

Oy		202 agctaacctgcgatgacccgcctctgcgcag-----cagtgtgaaaaagg	247
Db	121 AGCTAACCTGCGCATGCACCGCCTCTGGCGAGCAGGAGTGTGAACCAAGTTGTAATAAAGG	180	

**OY**    248 ggagaaagccactgtagcgcctgtactactaaaggccctggagaagtltcgatcatgtactaac 307  
|||||  
|||||  
**Db**    181 GGAGAGCCACTGGACGCTGTCACTGAAGGCCCTGGAGSGTTTCATCATGTACTAAC 240  
|||||  
|||||

QY 308 gccgaaggagacatgcttaccgtgcgaaatgtcagcaagcacctyggcctcagtcag 367

Db 241 gccgaggagacatggccttacctgtcgaaaatctcagcaagcacctgggcctcagtcag 3000

```

Oy 368 tggacctctgttcctctccctgatacataaaccacactctgtgaacaaattctcttg 427
Db 301 -----CTGG 304

```

QY 428 agccatctgacacaglatcttgatttatccacccctgaccaaagaactcaag 487  
|||||  
Db 305 AGGCATTTGACACACGATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAACTTCAAG 364

[illegible]

Qy	548	ac	tt	tc	ct	gc	ga	tg	ag	ca	gc	tc	ac	ca	gc	ag	gg	cg	ca	gc	tc	ca	ac	tc	aa	gc	607	
Db	425	ac	tt	tt	tc	cc	gc	gc	ga	tg	ag	ca	gc	tc	ac	ca	gc	ag	gg	cg	ca	gc	tc	ca	ac	tc	aa	gc

QY 608 cggccacctggaagtgtctgcactgtccagacatatgggcctacaagcccccgtgac 667

Db 485 CGGCCACCTTGGAAGGTGCTGCATCTCTCAGGACATATGAAGGCCCTTACAAGCCCCCTTCAC 544

QY	668	agacttcacctgcgcggagagccctcgctccgagcctccctgcgaatgctgtgtattatc	727
db	545	AGACTTCCCTCCCGGGAGCCTCGCTCCGAGCCTCCCTCGCAATGCTGTGTCATCT	604

QY	728	gtgaagccatccccagctccc	750
DB	605	GTGAAGCCATCCCNFACCCAGCC	627

RESULT 2  
AT322407

LOCUS	461 bp	mRNA	linear	EST 23-DEC
DEFINITION	m19b08.y1 Soares mouse p3NM19.5 Mus musculus cDNA clone			
IMAGE:463959	5' similar to TR:008787 008787 ENDOTHELIAL PAS DC			
PROTEIN 1	: mRNA sequence.			

ACCESSION	AI322407	
VERSION	AI322407.1	GI:4056836
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
1 (bases 1 to 461)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Schellenberg, R., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

**JOURNAL COMMENT**  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

This clone is available royalty-free through LNL ; contact the IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:27775

Seq primer: -40RP from Glibco  
This read has been verified (found to hit its original self in the correct orientation)

```

high quality sequence comp. 475:
FEATURES
source      Location/Qualifiers
1. .461
/organism="Mus musculus

```

```

/ab_xref="taxon:10090"
/clone="IMAGE:463939"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"

```

```

//AD_NOST-UNRES (ampicillin resistant),
//note="Vector: pTR73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,

```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector

(Pharmacists). Library went through one round of normalization to a Cot = 5. Library constructed by Bentz Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT	53 d	143 c	130 g	87 e
ORIGIN				

Query match	41.76%	Score 401	DB 51	Length 401
Best Local Similarity	100.0%	Pred. No. 4.1e-48		
Matches 461; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

```
Dy      16 gcaatgagcgttgggcctgcacgcgcttgagtctgaaccacccagcttcgcgaaggagaactcg /5
          |||||
Db       1 gccatgccgtttgggctgcaggcgcgttagtgtccaaccccagctgcggaaagagagatttcg 60
```

dy 76 cggagacgagcccgacagccgacgacacgagagcgcctgcacacacgcgcac 135  
 |||||  
 Db 61 cggagacgagcccgacagccgacgagagcgcctgcacacacgcgcac 120

**DY** 136 acctcgtcccttgcgcgcgagtcacgcgcacctgacaagccatcatatgcctc 195  
|||||  
**Db** 121 ACCTGCCCCTTGGCGCGGCGGTCAAGCCACACTGACAAGGCCCTCATATGCCTC 186

```

Oy      196 acaatcagcttaacctgcgatgacccgcctctgcgcagcaggtgaaaaaaagggaagcc 255
        |||
Db      181 ACAATCAGTACTCTGCATGACCGCCCTTGCGCAGCAGGTGAATAAAAGGGGAGAACC 240

```

QY 256 actgagcgcctgctactaccggaagccctggaaggttcgctcactggtactaccgcgcgaaggg 315  
|||||  
Db 241 ACTGAGCGCTGTACTACGTGAAGGCCCTTGAGAGGTTTGTCTATGTTACTACACGGCGGAGAGG 300  
QY 316 agacatggtactactgtcgtgaagaatgtcagcaagcaacctggcctcagtcagtcgaactc 375  
|||||  
Db 301 AGACATGGCTTACCTGCTGGGAAATGTCAGCAACACCTGGGCGCTCACTGACGTGACCTC 360  
QY 376 tcttcctccctccctgatacaataaccacactcctggtactcaattctctctcgaagctcatt 435  
|||||  
Db 361 TGTTCCTCTCTCCCTGATACATTAACCCACCTCTGCTGTTACCAATTTCTCTCTGAGCTCAT 420  
QY 436 ggaacacgatactctgattatcatccatccctcctgacccaaga 476  
|||||  
Db 421 GGACACAGTATCTTGTGATTTATCCATCCCTGTGACCAAGA 461  
RESULT 3  
AL528423 905 bp mRNA linear EST 13-FEB-2001  
LOCUS AL528423 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC029YH22 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL528423 GI:12791916  
VERSION AL528423.1 GI:12791916  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 905)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 905  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DC029YH22"  
/clone\_1id="LTI\_NFL003\_NBC3"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact : Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

|||||  
Db 120 TCACAGCGCTGCCCTTGCAGCGGCGGCTGACGCCACCTGGACAGGCTCATATCATCG 179  
QY 192 cctcaaatcagctaccctgcgcatgacccgctctgcgc-----agcaggt 237  
|||||  
Db 180 CCTACCATCAGCTACACCTGCGCATGACCGCTTGCAGCGGAGGAGTGGAACAGAGT 239  
QY 238 ggaanaaggggagagaccactggaacgctgtacttgaaggccctggaagggttcgcat 297  
|||||  
Db 240 GGGAGCAGGGGAGAGACCACTGATGATGCTGTACTGTAAGGCGCTGGAGGGCTTGAT 299  
QY 298 gttactcacgcgcggaaggagacatgcttaccctgctcgaagaatgtcagcaagcaacctggg 357  
|||||  
Db 300 GGTGCTACCGCCGAGGAGACATGCTTACCTGTCGAGATGTCACAAACACCTGGG 359  
QY 358 cctcagtcagtggaacctctgttctcctccctgtatacaataaccacactcctcgttacaat 417  
|||||  
Db 360 CCTCAGTCAG----- 369  
QY 418 ttctcctggaagctatggagacagatattgtattatcatccctcctgtgaccaagag 477  
|||||  
Db 370 -----CTGGAGCTCATTTGGACACAGCATCTTTGATTTCATTCACCCCTGTGACCAAGAG 423  
QY 478 gaacttcaagacgcgcctgaccccccagcgcaacctgtcacaagaagaagctggaagcccca 537  
|||||  
Db 424 GAGCTTCAGGAGAGCGCCCTGACCCCCCAGCAGACCCCTGTCCAGAGAGAGGTGAGAGGCCCC 483  
QY 538 acaagagcgcactttccctcgtcgaatgaagagcaagcgtccaccagcagagggcgacgctc 597  
|||||  
Db 484 ACGGAGCGGCTCTCTCTTGGCATGTAAGATACATCACACCGCGGGGCGACCTC 543  
QY 598 aacctcaagcgcgccactgtgaagtgctgctgactcctaagacataagggccataag 657  
|||||  
Db 544 AACCTCAAGCGCGCCACCTGGAAGTGTGACTCTTGACATATAGGGCTCTACAG 603  
QY 658 cccctcagacgaacttccctcgtcggagagcctcgtcgcagcctccctcctcaatgctg 717  
|||||  
Db 604 CCACTGGCGAGACTTTCACAGCTGAGGAGCCCTGACTCAGAGCCCCCGCTGAGTGGCTG 663  
QY 718 gtgcttatctgtgaagcatcccc 741  
|||||  
Db 664 GTGCTCATCTGCGAASCATCCCC 687  
RESULT 4  
AL519496 856 bp mRNA linear EST 13-FEB-2001  
LOCUS AL519496 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB004Y122 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL519496  
VERSION AL519496.1 GI:12782989  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 856)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 856  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DB004Y122"  
/clone\_1id="LTI\_NFL004\_NBC2"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"



Db 398 CATGTGACACAGCANTCTTGAATTCATCCACCCCTGTGACCAAGAGGAGCTTCAGACGC 457  
OY 492 cctgaaccccccagccgaacctgtcaagaagaagctggaagcccaagaagcgccactt 551  
Db 458 CCTGACCCCCCAGCAGACCTCTCCAGAGAGAAAGTGAGGCCCCACAGGAGCGGCTT 517  
OY 552 ttcctctggaatgaagacagacagctcacccagagagcgagctgcaacctcaaaagcgag 611  
Db 518 CTCCTGTGGGAG 577  
OY 612 caactggaagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 671  
Db 578 CACCTGGAAGGTGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 637  
OY 672 ttcctctggaagacagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 731  
Db 638 TTCTCAGCTGGAGCCCTGACTCAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 697  
OY 732 agccatcccc 741  
Db 698 ASCATCCCC 707

RESULT 6  
Bg699633 767 bp mRNA linear EST 07-MAY-2001  
LOCUS 602679365F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4811824 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg699633  
VERSION Bg699633.1 GI:13968143  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 767)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihai  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM10703 row: b column: 17  
High quality sequence stop: 616.  
Location/Qualifiers  
1. 767  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4811824"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified  
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag  
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROP 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 156 a 259 c 242 g 110 t  
ORIGIN

Query Match 37.1%; Score 408.2; DB 10; Length 767;

Best Local Similarity 78.0%; Pred. No. 1.2e-41;  
Matches 587; Conservative 0; Mismatches 93; Indels 73; Gaps 5;  
OY 6 cggcaagagggcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 65  
Db 14 CGACTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 73  
OY 66 ggaagaatcgcggagc 125  
Db 74 GGAAGAGTCCCGGAGATGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 133  
OY 126 gcttgccacactctgccccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 185  
Db 134 GGTAGCTACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193  
OY 186 catgcccacaaatagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 253  
Db 194 CATGCCCTCACCATTGAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 253  
OY 232 gaagtggaagaaaggag 291  
Db 254 CCAAGTGGAG 313  
OY 292 cgtcaatgctacacccgc 351  
Db 314 CGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 373  
OY 352 cctggcctcagtcagtgagcctgctgctgctgctgctgctgctgctgctgctgctgct 411  
Db 374 CTTGGGCTCTGAGTCAG----- 389  
OY 412 accaatctctcctgagagctgagacagatctgtatttaccctcctgtagc 471  
Db 390 -----CTGAGCTCATTTGGACACAGATCTTGATTTCAACCCCTGTGAC 437  
OY 472 caagagaaacttcaagaagccctgagcccaagccgaacctgtcaagaagaagctgga 531  
Db 438 CAAGAGAGACTTCAGAGACCCCTCGAACCCTGCAAGACCTGTCAGAGAGAGAGAGAG 497  
OY 532 gcccaacaagagcgccacttccctcggaatgaag-agaagcctacag-agaagagc 589  
Db 498 GCCCGCAGCAG 557  
OY 590 gaagctaacctcaagaagcgccactggaagtgctgctgctgctgctgctgctgctgctg 649  
Db 558 GCATCCTCAACCTCAAGGCGGAGACCTGGAAGGCTGTAAGCTGTAAGCTGTAAGCTG 617  
OY 650 cctacaagccccctgacagagcttc-ccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 708  
Db 618 CTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 677  
OY 709 caatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 741  
Db 678 CAGTGCCTGGCGCGCTCTGCGAAGCAATCCG 710

RESULT 7  
AUI18510 756 bp mRNA linear EST 19-OCT-2000  
LOCUS AUI18510 HEMBA1 Homo sapiens cDNA clone HEMBA1003760 5', mRNA  
DEFINITION sequence.  
ACCESSION AUI18510  
VERSION AUI18510.1 GI:10933569  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 756)  
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
Yamamoto,Y., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
Isigai,T.  
TITLE HRI human cDNA project



|||||  
Db 118 ACTGCGCCCTTGGCGGGGGGTCA-CACMNTNCTGACAAAGGCGCTCATCATATGGCCCTC 176  
|||  
Oy 136 acaatgaactactctgcatgcacacgcctctgcatgacagatggaagagggagagcc 255  
|||  
Db 177 ACAATGACACTACCTGCGCATG-ACGCGCTCTGCGCACAGANTGGAAGGAGGAGAGAC 235  
|||  
Oy 256 actgaagccctgcttaacctgaagggcggttctgcatgtaacctacacgcgcgaagg 315  
|||||  
Db 236 ACTGAGACCCCTGCTACCTGGAAGGCCCTGAGGGTTCTCATGACTACACCGCGAGGG 295  
|||  
Oy 316 agacatgctactactctggaagaaatgcatgaagacacctggccctgaatgagctc 375  
|||||  
Db 296 AGACATGCTTACCTGCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355  
|||  
Oy 376 tctctctctctctgatacaaacaccacccctctgtaaccaattctctctgagctcalt 435  
|||||  
Db 356 TGTTCCTCTCTCTGATACATTAACCCCATCTCTGCTACCAATTTCTCTGAGCTCATTT 415  
|||  
Oy 436 ggaacacagatcttgaatttatccatccctgctgtaaccaagaggaac 481  
|||||  
Db 416 GGACACAGTATCTTGAATTTAT-CATCCCTGTGACCAAGAGGAAAC 460  
|||

RESULT 9 633 bp mRNA linear EST 25-SEP-2001  
B1772109  
LOCUS 603059192P1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5208622 5',  
DEFINITION mRNA sequence.  
ACCESSION B1772109  
VERSION B1772109.1 GI:15763687  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 633)  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LHAM1523 row: 9 column: 23  
High quality sequence stop: 631.  
Location/Qualifiers

FEATURES  
Source  
1..633  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5208622"  
/clone\_lib="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source  
anonymous pool of 24 week female lung, 16 week female  
spleen, and 20-22 week male spleens. Library is oligo-dT  
primed and directionally cloned (Rcory site is destroyed  
upon cloning). Average insert size 1.4 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 026. Note:  
this is a NIH\_MGC library."

BASE COUNT 119 a 225 c 189 g 100 t  
ORIGIN

Query Match

30.5%; Score 336; DB 10; Length 633;

Best Local Similarity 79.2%; Pred. No. 9.7e-33;  
Matches 456; Conservative 0; Mismatches 50; Indels 70; Gaps 2;

Oy 44 ggtcgaacacccagctcgcgaaggaagatgcggygaagcggcccgacggcgagcc 103  
|||  
Db 114 GGTTCACACGACGAGTGTGGCAGGAAGAAAGTCCGGGATGCGGCCGACCCGCGCAGCC 173  
|||  
Oy 104 aggaagcggaggtgctgtacacagctcgcacacccctgcttgcgcgcgcgcgcgcgc 163  
|||||  
Db 174 AGGAGACCGAGGTGTGTGACAGCTGCTGCTACACGCTGCTTCCGCCGCGCTACGCG 233  
|||  
Oy 164 cgcacctggaacagcctccatcatgctgcctcacatcagctactgcatgacacggcc 223  
|||||  
Db 234 CCCACCTGGACAAAGGCTCTATCATGCGCTCACCATGATGATGATGATGATGATGATG 293  
|||  
Oy 224 tctgcgc-----agcaggtggaagaaaggggagagccactgtagcgcctct 269  
|||||  
Db 294 TGTGCGCGCGACAGGAGAGTGGAAACAGTGGAGCAGAGGGGAGAACCACTGATGCTGCT 353  
|||  
Oy 270 acctgaagcctctggaaggttctgcatgtaactaccgcgcgcgcgcgcgcgcgcgcgc 329  
|||||  
Db 354 ACCTGAAGGCCCTGGAAGGCTTGTCTATGCTCTACCGCCGAGGAGACATGGCTTACC 413  
|||  
Oy 330 tctcgaagaaatgcatgaagacacctgggcctcagtcagtgagcctctgtctcctccct 389  
|||||  
Db 414 TGTGCGAGAGATGTCAAGAAACACCTGGGCTCAGTACG----- 451  
|||  
Oy 390 gatacataaccac 449  
|||||  
Db 452 -----CTGAGCTCATTTGACACAGCATCTT 477  
|||  
Oy 450 tgatttatccatccctgtagacaagaaggaactcaagacgcctgaccccgagcgga 509  
|||||  
Db 478 TGAATTCATCCACCCCTGTGACCAAGAGAGCTTCAGAGCGCCCTGACCCCGACAGAC 537  
|||  
Oy 510 ccttcgaagaaagatgtagaagccccaagacgcgcacatttccctggaatgaagag 569  
|||||  
Db 538 CCGTGCAGAGAGGAAGGTGAGAGGCCCGCCAGAGGCGGTCTCTCTCGCATGAAGAG 597  
|||  
Oy 570 cagcgtacacagcagagagggcgacgctcaactcaa 605  
|||||  
Db 598 TACGCTACACGACCGCGGCGACACCTCAACCTCAA 633  
|||

RESULT 10 459 bp mRNA linear EST 03-SEP-1996  
AA041707  
LOCUS m102g02.r1 Soares mouse embryo NDME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:474962 5', mRNA sequence.  
ACCESSION AA041707  
VERSION AA041707.1 GI:1519863  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 459)  
Marrin,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kuwada,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterson,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.





QY	1034	casg	1036
Db	3	CAG	1
RESULT	12		
LOCUS	BM118909/c		
DEFINITION	LM917H06-3 N1A Mouse Newborn Kidney cDNA Library (Long)	425 bp	mus
ACCESSION	BM118909		
VERSION	BM118909.1	GI:17086935	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 425)		
AUTHORS	Piao, Y., Karpiul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K. and Ko, M.S.H.		
TITLE	Systematic Analyses of N1A Mouse Newborn Kidney cDNA Library		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA Email: cdna@igsun.grc.nia.nih.gov Plate: LM917 row: H column: 06 Seq primer: -21M13 Forward High quality sequence stop: 425 POLYA=yes.		
FEATURES	Location/Qualifiers		
SOURCE	1..425		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="niaEST:LM917H06-3"		
	/db_xref="taxon:10090"		
	/clone="LM917H06"		
	/clone_lib="N1A Mouse Newborn Kidney cDNA Library (Long)"		
	/tissue.type="Newborn kidney"		
	/dex_stage="Newborn"		
	/lab_host="DH10B"		
	/note="Organ: kidney; Vector: pSPORT1 (Invitrogen); Site_1: Salt; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long transcript enriched cDNA library (Ref. Genome Res. 11:1553-1558 (2001)). [PMID:11544191]]. In brief ,double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-		
	pgactactgtcagatgcgcagcggccccccttttttttttt-3'] from 26 microgram of total RNA, treated with R4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Salt4, purified by phenol/chloroform, and separated from free linkers by centrifugation 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Salt-S. The products were purified by phenol/chloroform and Gentricon 100. The cDNAs were digested with Salt and NotI enzymes, and cloned into Salt/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao (NIA)."		
BASE COUNT	95 a	96 c	136 g
ORIGIN	27.1%; Score 298.2; DB 10; Length 425; Best Local Similarity 99.0%; Pred. No. 4.8e-28; Matches 300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		

QY	734	ccatcccccaagtcctccctcccaacgagtggtgctactcttggtctctccacaggaagaaagtc	793
Db	303	CTAACCCCCAGCTCCCCCTTCCACGATGGGTACTCTGGGTCTTCCACACGAAAGACTC	244
QY	794	ccatctctacattatccaccctctcttggaagcactactcttgctgtcaagaagtgagc	853
Db	243	CCATGCTCTACCTTATTCACCCCTCTTTGGAAAGGCACTACTTGTCTTGTCAAGAGGTGGC	184
QY	854	ctgttcaggctgtctacaggggaaagggactgaaatcctctctccctaatgggtgtgtggg	913
Db	183	CTGTTCAGGTGCTACAGGGGAAAGGACATGAATCTCTCCCTCATGGGTGTGGGG	124
QY	914	cccttaaacgggaaaattctctgtgcttaaggaaggtgaagacatggcccaactatcct	973
Db	123	CCCTTACCCGAAATAATGTCTCTGGCTAGAGGGAGGAAAGGACATGGCCACGTTACTCT	64
QY	974	tagccacagaacccacaataatgtctccaaaacccacataaagacctctccttgttaagcac	1033
Db	63	TAGCCAGAAACCCACAAATGTCTCCAAACACCATTAAGACCTCTCCTTGTAGGCAC	4
QY	1034	cag 1036	
Db	3	CAG 1	
RESULT 13			
LOCUS	BM119659	610 bp	mRNA linear EST 01-FEB-2000
DEFINITION	L0929A01-3 NIA Mouse Newborn Kidney cDNA Library (Long)	Mus	
ACCESSION	musculus cDNA clone L0929A01 3', mRNA sequence.		
VERSION	BM119659		
KEYWORDS	BM119659.1 GI:17087685		
SOURCE	EST		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 610)		
AUTHORS	Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Lim,M.K. and Ko,M.S.H.		
TITLE	Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Dawood B. Dudekula		
	National Institute on Aging/National Institutes of Health		
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA		
	Email: cdna@jgsum.gic.nia.nih.gov		
	Plate: L0929 row: A column: 01		
	Seq primer: -21M13 Forward		
	High quality sequence stop: 610		
	POLYA=yes.		
FEATURES			
Source	Location/Qualifiers		
	1..610		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="naest:L0929A01-3"		
	/db_xref="taxon:10090"		
	/clone="l0929A01"		
	/clone_1lb="NIA Mouse Newborn Kidney cDNA Library (Long)"		
	/tissue_type="Newborn Kidney"		
	/dev_stage="Newborn"		
	/lab_host="DH10B"		
	/note="Organ: Kidney; Vector: pSPORT1 (Invitrogen);		
	Site:1: Salt; Site:2: NotI: Mouse cDNA project by the		
	Laboratory of Genetics, National Institute on Aging (NIA)		
	Intramural Research Program, NIH		
	(http://jgsum.gic.nia.nih.gov/cDNA). This is a		
	long transcript enriched cDNA library (Ref: Genome Res.		
	11:1553-1558 (2001). [PMID:11544199]). In brief		
	, double-stranded cDNAs were synthesized with an Oligo(dT)		
	primer (Invitrogen: 5'-		
	pgactaggtcttgatcgcgacaggcccccctttttttttttt-3') from 26		



TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp

HRI human cDNA project: 5'-6 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source location/Qualifiers

1..873  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEMBA1000216"  
/clone\_lib="HEMBA1"  
/tissue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/note="Vector: pME18SFL3"

BASE COUNT 175 a 305 c 248 g 142 t 3 others  
ORIGIN

Query Match 23.0%; Score 253.2; DB 9; Length 873;  
Best Local Similarity 86.6%; Pred. No. 1.5e-22;

Matches 279; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 424 ctggagctcatggaacagatatttgcattatccatccctgtgaccaagaagaatt 483  
DB 281 ctggagctcatggaacagatatttgcattatccatccctgtgaccaagaagaatt 483  
OY 484 caagaagccctgaaccccaagcgaacctgtcaaaagaagaagctggaagccccaacagag 543  
DB 341 CAGGACGCCCTGAGCCCGCAGACAGCCTGTCAGAGGAGAGGTGAGGCGCCCGCAGGAG 400  
OY 544 cgcaccttccctcgaatgaagaagcgtccaccagagaagggcgaagctcaactc 603  
DB 401 CGGTCTCTCTCTGCGCATGAAGATGACGTCACACGCCGCGGCGCACCTCAACCTC 460  
OY 604 aaagcggcaacctggaaggtgtgcacgtcgaagacatagaaggcctacaagccccc 663  
DB 461 AAGCGGCCACCTGGAAGGTGCTGAAGTGTCTGGAATATGAGGGCTTACAAGCCACCT 520  
OY 664 gcaacagacttccctcgaagcgtcgaagcgtccaccctcgaagcgtcgaagcgtcgt 723  
DB 521 GCGCAGACTTCTCCAGCTGGAGCCCTGACTCAGAGCCCGCGCTGAGTGCCTGCTGCTC 580  
OY 724 atctgtgaagcattccccagc 745  
DB 581 ATCTGGAAGCATCCCCACC 602

Search completed: September 8, 2002, 05:57:29  
Job time: 6842 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 05:34:17 : Search time 2212.99 Seconds  
(without alignments)  
10401.850 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 1100  
Sequence: 1 gaattcgccagcgagccat.....aaaaaacatgcgccgc 1100

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query Match Length DB ID Description  
No. Score Match Length DB ID Description

1	1032	93.8	1032	10	AF416641	AF416641 Mus muscu
2	556.6	50.6	2207	10	AF060194	AF060194 Mus muscu
3	505.4	45.9	2077	10	RN0277827	AJ277827 Rattus no
4	471.6	42.9	2082	9	AK027725	AK027725 Homo sapi
5	469.4	42.7	2023	9	AF463492	AF463492 Homo sapi
6	451	41.0	2004	9	AB054067	AB054067 Homo sapi
7	447.4	40.7	2595	9	AK021653	AK021653 Homo sapi
8	301.6	27.4	250615	2	AC073814	AC073814 Mus muscu
9	276	25.1	134189	2	AC079550	AC079550 Mus muscu
10	265.4	24.1	345	9	AF079154	AF079154 Mus muscu
11	253.4	23.0	1708	9	AK021421	AK021421 Homo sapi
12	228.2	20.7	167150	2	AC095259	AC095259 Rattus no
13	211.6	19.2	1841	9	AK024095	AK024095 Homo sapi
14	196	17.8	134189	2	AC079550	AC079550 Mus muscu
15	192	17.5	3605	5	AF304864	AF304864 Oncorhinc
16	191	17.4	191	10	MMH1F3A02	AF079141 Mus muscu
17	176.8	16.1	2888	4	AB018399	AB018399 Bos tauru
18	174.6	15.9	2730	9	HS051626	U51626 Human MOP2
19	173	15.7	2816	6	I84559	I84559 Sequence 1
20	173	15.7	2818	9	HS081984	U81984 Human endot
21	173	15.7	146180	9	AC007193	AC007193 Homo sapi
22	172	15.6	2776	6	AX330581	AX330581 Sequence
23	169.6	15.4	3488	10	D89787	D89787 Mus musculu
24	168	15.3	3415	10	AF045160	AF045160 Mus muscu
25	160	14.5	2730	10	RN0277828	AJ277828 Rattus no
26	156.6	14.2	3031	6	I84560	I84560 Sequence 2
27	156.6	14.2	3033	10	MMU081983	U81983 Mus musculu
28	148.8	13.5	6597	5	AB013746	AB013746 Gallus ga
29	146.2	13.3	3718	10	AF057308	AF057308 Rattus no
30	144.6	13.1	2711	10	RNHYPFAC1	Y09507 R. norvegicu
31	139.8	12.7	3973	10	AF003695	AF003695 Mus muscu
32	138.2	12.6	3746	6	AX306008	AX306008 Sequence
33	138.2	12.6	3746	10	MMH1F1ALP	X9580 M. musculus
34	136.4	12.4	2640	5	AF129813	AF129813 Gallus ga
35	136.2	12.4	2551	4	AB018398	AB018398 Bos tauru
36	135.8	12.3	146	10	MMH1F3A03	AF079142 Mus muscu
37	135.4	12.3	3076	5	AF212989	AF212989 Coturnix
38	135	12.3	3867	10	MMU59496	U59496 Mus musculu
39	134.6	12.2	2509	9	AF304431	AF304431 Homo sapi
40	134.6	12.2	2522	9	AF208487	AF208487 Homo sapi
41	134.6	12.2	2528	6	AX230580	AX230580 Sequence
42	134.6	12.2	2530	9	AF207601	AF207601 Homo sapi
43	134.6	12.2	2537	9	AF207602	AF207602 Homo sapi
44	134.6	12.2	3551	9	AB073325	AB073325 Homo sapi
45	134.6	12.2	3678	9	HSU22431	U22431 Human hypox

## ALIGNMENTS

RESULT 1  
AF416641  
LOCUS AF416641 1032 bp mRNA linear ROD 09-DEC-2001  
DEFINITION Mus musculus inhibitory PAS domain protein (Ipas) mRNA, complete  
cds.  
ACCESSION AF416641  
VERSION AF416641.1 GI:17432318  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 1032)  
Makino,Y., Cao,R., Svensson,K., Bertilsson,G., Asman,M., Tanaka,H.,  
Cao,Y., Berkenstam,A. and Poellinger,L.  
Inhibitory PAS domain protein is a negative regulator of  
hypoxia-inducible gene expression  
Nature 414 (6863), 550-554 (2001)  
JOURNAL PUBMED  
11734856  
REFERENCE  
2 (bases 1 to 1032)  
Makino,Y., Cao,R., Svensson,K., Bertilsson,G., Asman,M., Tanaka,H.,  
Cao,Y., Berkenstam,A. and Poellinger,L.  
Direct Submission



RPPVARRPRANSFHGLSPIDPELIPRMGSDPRLNCSSPSKGRDPTSLMPGTRKR  
ALAOSEDKGLLELETKPKRSPRLPGSFLPLSLFLQROLGNOQDPRALV  
HSHPEPLGLPSLSTLCQHEETOPRHHFPFPAAGLQGTN"  
BASE COUNT 471 a 742 c 597 g 397 t  
ORIGIN

Query Match 50.6%; Score 556.6; DB 10; Length 2207;  
Best Local Similarity 88.1%; Pred. No. 1.1e-102;  
Matches 658; Conservative 0; Mismatches 19; Indels 70; Gaps 2;

QY 26 tggggtcgcagcgcgtgagtcgaaacacgcagctctgcgaaggaagtcgcggaacgcg 85  
DB 85 TGGAGCTGGGACCAAGACAGTCCAGACCGAGCTCGGGAAGGAATCCCGGAGCCGG 144  
QY 86 cccgcagc 145  
DB 145 CCGGACAGCCGGCCGACGACGAGAGCGAGGTGCTGACAGCTGGCGCACACTCGCCCT 204  
QY 146 ttgc 205  
DB 205 TTGCGGCGCGGCGTACGCGCGCCACCTGACCAAGCCCTTCATCATGCGCTCAATCAGCT 264  
QY 206 acctgc 251  
DB 265 ACCTGGCGCATGCACCGCGCTTCGCGACGAGGAGTGAACAGTGAAGGAAAGGAGGAG 324  
QY 252 agcgaactgc 311  
DB 325 AGCCACTGGAGCGCGCTGCTACCTGAAGCCCTTGAGAGGCTTGCATGCTACTACCGCCG 384  
QY 312 agggagacacgc 371  
DB 385 AGGAGACATGCTTACCTCTCGAATAATGACCAAGCAACCTGGCGCTCAAGTACG--- 440  
QY 372 cctctgtct 431  
DB 441 -----CTGGAGCT 448  
QY 432 catgtgaacacgatacttttaattatcatccctgtgacaaaggaagacttcaagaagc 491  
DB 449 CATTGGACACAGTATCTTTTATTTTATTCCTGACCAAGAGCAACTTCAAGACGC 508  
QY 492 cctgaaccccgagc 551  
DB 509 CCGTACCCCGGAGGCGGACCTGTCAAGAGAGAGCTGGAAGCCCAACAGAGCGCCACTT 568  
QY 552 ttccctgcgaatgaagagc 611  
DB 569 TTCCCTGCGAATGAGAGGACGCTCACACGACGAGGCGGACGCTCAACCTCAAAAGCGCG 628  
QY 612 caectggaaggtgc 671  
DB 629 CACCTGGAAGGTCTGCTGACCTGCTACGACATATGAGGCGCTTACAAAGCCCTTGACACGAC 688  
QY 672 ttccctgc 731  
DB 689 TTCCCTGCGGAGGAGCGCTGCTGCTGAGCTCCCTGCAATGCTGCTGCTGCTGCTG 748  
QY 732 agccatcccccaagctccctctcaca 758  
DB 749 AGCCATCCCGGACGACGACGAGTCTGGA 775

RESULT 3  
RNO277827 2077 bp mRNA linear ROD 26-FEB-2001  
LOCUS Rattus norvegicus mRNA for hypoxia inducible factor 3 alpha (Hif-3a  
DEFINITION gene).  
ACCESSION AJ227827  
VERSION AJ227827.1 GI:8953574  
KEYWORDS Hif-3a gene; hypoxia inducible factor 3 alpha.  
SOURCE Norway rat.

ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2077)  
AUTHORS Kietzmann,T., Cornesse,Y., Brechtel,K., Modaresi,S. and  
Jungermann,K.  
TITLE perivenous expression of the mRNA of the three hypoxia-inducible  
factor alpha-subunits, HIF1alpha, HIF2alpha and HIF3alpha, in rat  
liver  
JOURNAL Biochem. J. 354 (Pt 3), 531-537 (2001)  
MEDLINE 21134367  
REFERENCE 2 (bases 1 to 2077)  
AUTHORS Kietzmann,T.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAY-2000) Kietzmann T., Biochemistry I, Institut of  
Biochemistry and Molecular Cell Biology, Humboldtallee 23, D-37073  
Goettingen, GERMANY

FEATURES  
source location/Qualifiers  
1..2077  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/sex="male"  
/cell\_type="primary hepatocyte"  
/tissue\_type="liver"  
/dev\_stage="adult"  
/country="Germany"  
5..1993  
/gene="Hif-3a"  
5..1993  
/function="transcription factor"  
/codon\_start=1  
/product="hypoxia inducible factor 3 alpha"  
/protein\_id="CAB96611.1"  
/db\_xref="GI:8953575"  
/db\_xref="SPRMBL:09JHS2"  
/translation="MDMDDRSSTELRKRSRPARSRSEMEVLYXLAHLPFARG  
VSANHDKASIMRLTITLYLRNHRICAGENKVRKEBPIDACTLKLBGFVYVLAG  
DMAYLSENVSKHIGLSLELIGHSITDFIPPCDELDLALTPRSLSKKSSAATRG  
HFSLRKSTLYSRGALNKAATWKVLIHSGHRAVKPPAOTSPASPSPEPLQCLV  
LICEAIPHPASLEPPLRGAFLSRHSIDMKFTYCDERLAEVAGSPDDIGCSAYEI  
HADDPSVRSITLISKGOAVYGOVRFPLARTGYDGTQATVAVVSGGPOSEITIC  
VHFLISRVENSGVILSLDTEQTRRPOLGTSKKGITQONSIDVAPRPLATLHPRA  
LSEASLAADPRFRRCSPDLRLMAPITIDEPPTATPTPTQAAKRPQSPPLADLPDQALV  
GLENAHRLSTARNTMETDLDLADQDPDPLDLEMLAPYLSMDDFQINSSEQLPKVHR  
RPPRTARRRARRSFHGLSPIDPEALPLPRMGSDPRLNCSSPSKGRDPTALPRTKR  
ALAOSEDKGLLELETKPKRSPRLPGSFLPLSLFLQROLGNOQDPRALV  
DSHEPPLGLPSLSTLCQHEETIGSRNHPPLAAGLQGTN"

BASE COUNT 457 a 700 c 552 g 368 t  
ORIGIN

Query Match 45.9%; Score 505.4; DB 10; Length 2077;  
Best Local Similarity 84.3%; Pred. No. 2.5e-92;  
Matches 623; Conservative 0; Mismatches 46; Indels 70; Gaps 2;

QY 26 tggggtcgcagcgcgtgagtcgaaacacgcagctctgcgaaggaagtcgcggaacgcg 85  
DB 6 TGGAGCTGGGACCAAGACAGTCCAGACCGAGCTCGGGAAGGAAGTCCCGGATCCG 65  
QY 86 cccgcagc 145  
DB 66 CCGGACAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 125  
QY 146 ttgc 205  
DB 126 TTGCGGCGCGGCGTACGCGCACACTGACCAAGGCTCCATCATGCTGCTGCTGCTGCTG 185  
QY 206 acctgc 251  
DB 186 ACTTGGCGCATGCACCGCGCTGCGCTGACGAGGAGTGAACAGGAGTGAAGAAAGAGGAGG 245

[illegible]

			/organism="Homo sapiens"
		/db_xref="taxon:9606"	
		/clone="OVARC1000241"	
		/tissue_type="ovary, tumor tissue"	
		/clone_1fb="OVARC1"	
CDS		/note="Cloning vector: pMEI8SFLJ3"	
		13..1911	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAB55324.1"	
		/db_xref="GI:14042618"	
		/translation="MALGLOARSTTELRKESRDARRSRQETEVLYOLAHPLPRA RGVNAIHDKSIIMRLTISYLEMRHLCAAGENYQAGGEPIADALNCTALBDFWVLNR EGDAVISENVSKHLGISOLELIGHSITFDLIHPDQDELIDALTPOOTLSRRVEAF ERCSYLNRKSTLITSRGRFLNLKAATWKLVLNCSGHMRYKPPVOYSPGSPEEPLOOC LVLLCEAIIPHPGSLIEPLRGCAFLESHSLDMKEFYCDRIAEVAGSPDDILGCSAYE IVHALSDAVKSIIHTLLSKGAQTGYRFLEARGGVLMTOATVAVSGSGPOSESI VCVFELISOYEERFENVLSIRQTEBOHSRRPILORGA.POKDKPNPDSDLDTPGPRILARL HPSELSEPALAADPRRCSPDIRLRILPILIDGNASVAATPSPILATRHROPISLADLPD ELPVGTENVHRLFTSGKDTEAVENTDLIDADDALDIEMLAPLYISGDIDPOLINSBED PRAHRPLGAVPRPARASFGLSPALPELSPILPMWGSDPR.LSCSSPARGPSASSPMAA GARRTTLAOSSEDEDEGVELLGVPRPRSPHEHENLPLSLVCGINILIMPSTLP SWLKPYLV"	
BASE COUNT	416 a	684 c	607 g    375 t
ORIGIN			
Query Match	42.9%; Score 471.6; DB 9; Length 2082;		
Best Local Similarity	79.9%; Pred. No. 1.7e-85;		
Matches 612; Conservative	0; Mismatches 84; Indels 70; Gaps 2;		
Oy	7	ggcacgagggccatgctgtgggagtgcagcgctgaagtcgnacaccgagctgcggaag	66
Dd	1	GACGTGCAGCCATGGCGCTGGGAGCTGCAGCGCGCAAGTCGACCAGCGAGCTGGCGCAAG	60
Oy	67	gagaagtcgcgggagcgcgccgcagccgcgcgcaagcaggaagcggaggctgtactaacg	126
Dd	61	GAAAAGTCCCGGAGAGCGGCCGCCGCGCGGCGCACCCAGGAGACCGAGAGTGCTGTATCACG	120
Oy	127	ctggggacacctctgccccttttgccggggggttaacggcgacaccttggaaagaagccccaat	186
Dd	121	CTGGCTTCACAGCTCTCCCTTGCCCCGGCGGTCAACGCCACCTGGACAAGGCCCTCTATTC	180
Oy	187	atgcgctacaatatgaactgactctgcgcatgaacgcgcctctgcgc-----ag	232
Dd	181	ATGGCCCTCACCATAGCTACTCTGGCATGCGACCGCCCTCTGCCGCCGCAAGGGGAGTGGAAC	240
Oy	233	caggttgaaaaaagggggagagccacttgaagcgcctgtactacttgaagcccttgaaggttc	292
Dd	241	CAGGTGGGAGGACGAGGGGGAACCACTGCGATGCTCTCTACTCGAAGGCCCTGGAGGGCTTC	300
Oy	293	gtcatagtactcacccgcgcgaaggagacatgtcttaactgtctggaanaatgttaagaagac	352
Dd	301	GTCAATGTTGCTACCGCCGAGGAGAACATATGGCTTAACCTGTGCGGAAGAATTGACGAACAC	360
Oy	353	ctgggcccactgacatgtgaactctgtcttcctccctatacataaaccacactccgtgta	412
Dd	361	CTGGGCTTCAGTCAAG-----	375
Oy	413	ccaattctctctggagactatlggacacagtatcattgatattatcatccctgtgacc	472
Dd	376	-----CTGAGAGCTCATTTGGACACACACATCTTTGATTTCATCACCCCTGTGACC	424
Oy	473	aagaggaactccaagaagcctctgaaccccccgagccggaactcttcaaagaagaagcttgaag	532
Dd	425	AAGAGGACTTTCAGAGACCCCTTGACCCCACGAGACCTTTCACAGGAGGAAGTGTGAGG	484
Oy	533	ccccaaacagagcgcaactcttccctctgcgaatgaagaagacgctaccaagcaggaagcgaca	592
Dd	485	CCCCACGAGACGGGTCTTCTCTCTTGGCCATGAAGAGTAGTACCTCACACGCCGCGAGCGCA	544
Oy	593	cgtcaaacctaagaagcgccacacttgaaggtgtcttcaactgtctcaggaacatatgagggct	652



[illegible]

OY	552	acgtccaaaccttaagaagcgaccacccggaagtgcctgcactctcaaggacatatagggcc	651
Db	558	ACCTCTAACCTTCMAAGGCGGCCACTGTGAAGGTGCTGCATGCTCTGGACATATGAGGGCC	617
OY	652	tacaagccccctgcacagacttcctcctgcgcggaagcctgcctccgaagcccccctgcaa	711
Db	618	TACAAGCCACCTCGCCGACACTTCTCAGCTGGGAGCCCTTGACTACGAGACCCCCGCTGAG	677
OY	712	tgcctgtgcttatctgtcaagccalcgccccaagctcccttccaccga	758
Db	678	TGCCGTGGTCTCATCTGGAGAAGCCATCCCACCACCAAGGAGAGCTTGA	724
RESULT	6		
AB054067			
LOCUS			
DEFINITION	AB054067	2004 bp mRNA linear PRI 18-JAN-2002	
	Homo sapiens HIF-3A mRNA for hypoxia-inducible factor-3 alpha,		
	complete cds.		
ACCESSION	AB054067		
VERSION	AB054067.1	GI:16040974	
KEYWORDS			
SOURCE			
ORGANISM		Homo sapiens kidney cDNA to mRNA.	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE			
AUTHORS		Hara,S., Hamada,J., Kobayashi,C., Kondo,Y. and Imura,N.	

**TITLE** Expression and characterization of hypoxia-inducible factor (HIF)-3 $\alpha$  in human kidney: suppression of HIF-mediated gene expression by HIF-3 $\alpha$

**JOURNAL MEDLINE** Biochem. Biophys. Res. Commun. 287 (4), 808-813 (2001)

**REFERENCE** 2145827

**AUTHORS** 2 (bases to 2004)

**TITLE** Hara, S., Kobayashi, C., Hamada, J., and Imura, N.

**JOURNAL** Direct Submission

Submitted (15-JAN-2001) Shuntaro Hara, Kitasato University, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan (E-mail: harsaspharm.kitasato-u.ac.jp, Tel:81-3-5791-6265, Fax:81-3-3442-4146)







```
gene      <1. .>345
          /gene="HIF3A"
CDS       <1. .>345
          /gene="HIF3A"
          /codon_start=1
```

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Heliix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science,

```

FEATURES             University of Tokyo.
Source               Location/Qualifiers
                    1..1708
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="HEMBA1000216"
                    /tissue_type="whole embryo, mainly head"
                    /clone_lib="HEMBA1"
                    /dev_stage="embryo, 10 weeks"
                    /note="cloning vector: pME18SFL3"
                    125..724
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="BAB13819.1"
                    /db_xref="GI:10432601"
                    /translation="MRPAAAGARPRCCTSMLTRCPSPASAPMTWRPLSCASPSATC
                    ACTASAPOLIELGHSIRPDIFHPCDEEODALTPOOTLSRKVEAPTRCSTLRKST
                    LTRSRRTNLKRAATKYNCSGHHARAYPPNOTSPAGSPDEPPLQCLVLCEALPH
                    GSLEPPLGRGAFSLSHSDMKFTYCDNVAEAGVSPMT"
                    280..281
                    /note="146 bases segment is present in AK21653, probably
                    caused by alternative splicing."
                    misc_difference 714..715
                    /note="compared to AK21653 and AC007193.1"
                    /replace="c"
BASE COUNT          340 a          588 c          461 g          319 t
ORIGIN
Query Match          23.0%; Score 253.4; DB 9; Length 1708;
Best local Similarity 84.8%; Pred. No. 2.6e-41;
Matches 284; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 424 ctggagcattggagacagatatttgcattatccatccctgtgaccagaagaatt 483
    |||||
Db 281 ctggagcattggagacagacatctttgatttcatccaccctgtgaccagaagaatt 340
    |||||

Oy 484 caagacgcccctgacccccagcgaacctgtcaagaagaagctggaagcccaacagag 543
    |||||
Db 341 CAGGAGCGCCCTGACCCCGCAGAGACCTGTCCAGAGAGAGAGAGAGAGAGAGAGAG 400
    |||||

Oy 544 cgcacatttccctcgtgaatgaaagacagctccaccagagagggcgcagctcaactc 603
    |||||
Db 401 cggctccttcttcttgcgcatgaaagatgacacacgacgacgacgacgacgacgac 460
    |||||

Oy 604 aaagcggcaccctggaaggtgctgcacgctcagagacatagaggcccaaacccctc 663
    |||||
Db 461 AAGGCGGCGCACCCTGGAAGGTGCTGTAAGTCTGTGACATATAGAGGCGCTACAGCA 520
    |||||

Oy 664 ggcagacattccctcgcggagacccctcgcgcagcctccctcgaatgctgtgctt 723
    |||||
Db 521 GCGCAGACTTTCAGACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
    |||||

Oy 724 atctgtgaagcattcccccagctccctcgaaga 758
    |||||
Db 581 ATCTGTGAGAGCATTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
    |||||

RESULT 12
AC095259          167150 bp      DNA      linear      HTG 20-DEC-2001
LOCUS            Rattus norvegicus clone CH230-10B18, *** SEQUENCING IN PROGRESS
DEFINITION      *** 67 unordered pieces.
ACCESSION        AC095259
VERSION          AC095259.2 GI:17943467
KEYWORDS         HTG; HTGS_PHASE1.
SOURCE           Norway rat.
ORGANISM         Rattus norvegicus
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                 Rattus.
REFERENCE        1 (bases 1 to 167150)
AUTHORS          Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

```

```

TITLE
JOURNAL
AUTHORS
REFERENCE
JOURNAL
COMMENT

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Bowle,S., Briveva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Donthaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,O., King,L., Koryah,J.,
Kovari,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisged,R., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapus,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nmekenwo,S.,
Ogub,M., Okwoung,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Pines,E., Pul,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Woodson,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
2 (bases 1 to 167150)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625813.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDX
Center clone name: CH230-10B18
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findphraplist
Consensus quality: 128712 bases at least Q40
Consensus quality: 139536 bases at least Q30
Consensus quality: 148385 bases at least Q20
Estimated insert size: 136187; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 5997: contig of 5997 bp in length  
\* 5998  
\* 6098 6097: gap of unknown length  
\* 12971 12971: contig of 6874 bp in length  
\* 13072 13071: gap of unknown length  
\* 15335 15335: contig of 2264 bp in length  
\* 15336 15335: gap of unknown length  
\* 15436 21889: contig of 6454 bp in length  
\* 21890 21889: gap of unknown length  
\* 21990 26955: contig of 4966 bp in length  
\* 26956 27055: gap of unknown length  
\* 31737 31737: contig of 4682 bp in length  
\* 31738 31837: gap of unknown length  
\* 35633 35633: contig of 3796 bp in length  
\* 35634 35733: gap of unknown length  
\* 35734 38570: contig of 2837 bp in length  
\* 38571 38670: gap of unknown length  
\* 43208 43207: contig of 4537 bp in length  
\* 43308 48003: gap of unknown length  
\* 48004 48103: gap of unknown length  
\* 48104 50814: contig of 2711 bp in length  
\* 50815 50914: gap of unknown length  
\* 50915 54185: contig of 3271 bp in length  
\* 54186 54285: gap of unknown length  
\* 54286 58646: contig of 4361 bp in length  
\* 58647 58746: gap of unknown length  
\* 58747 61143: contig of 2397 bp in length  
\* 61144 61243: gap of unknown length  
\* 61244 64596: contig of 3353 bp in length  
\* 64597 64696: gap of unknown length  
\* 64697 67316: contig of 2620 bp in length  
\* 67317 67416: gap of unknown length  
\* 67417 70770: contig of 3354 bp in length  
\* 70771 70870: gap of unknown length  
\* 70871 73844: contig of 2974 bp in length  
\* 73845 73944: gap of unknown length  
\* 73945 77641: contig of 3697 bp in length  
\* 77642 77741: gap of unknown length  
\* 77742 79928: contig of 2187 bp in length  
\* 79929 80028: gap of unknown length  
\* 80029 82786: contig of 2758 bp in length  
\* 82787 82886: gap of unknown length  
\* 82887 84591: contig of 1705 bp in length  
\* 84592 84691: gap of unknown length  
\* 84692 87672: contig of 2981 bp in length  
\* 87673 87772: gap of unknown length  
\* 87773 90407: contig of 2635 bp in length  
\* 90408 90507: gap of unknown length  
\* 90508 93319: contig of 2812 bp in length  
\* 93320 93419: gap of unknown length  
\* 93420 96270: contig of 2851 bp in length  
\* 96271 96370: gap of unknown length  
\* 96371 98956: contig of 2586 bp in length  
\* 98957 99056: gap of unknown length  
\* 99057 100605: contig of 1549 bp in length  
\* 100606 100705: gap of unknown length  
\* 100706 104118: contig of 3413 bp in length  
\* 104119 104218: gap of unknown length  
\* 104219 105843: contig of 1625 bp in length  
\* 105844 105943: gap of unknown length  
\* 105944 108705: contig of 2762 bp in length  
\* 108706 108805: gap of unknown length  
\* 108806 110740: contig of 1935 bp in length  
\* 110741 110840: gap of unknown length  
\* 110841 112762: contig of 1922 bp in length  
\* 112763 112862: gap of unknown length  
\* 112863 115666: contig of 2804 bp in length  
\* 115667 115766: gap of unknown length  
\* 115767 117269: contig of 1503 bp in length  
\* 117270 117369: gap of unknown length  
\* 120158: contig of 2789 bp in length

\* 120159 120258: gap of unknown length  
\* 120259 122019: contig of 1761 bp in length  
\* 122020 122119: gap of unknown length  
\* 122120 124117: contig of 1998 bp in length  
\* 124118 124217: gap of unknown length  
\* 124218 126395: contig of 2178 bp in length  
\* 126396 126495: gap of unknown length  
\* 126496 127499: contig of 1004 bp in length  
\* 127500 127599: gap of unknown length  
\* 127600 129051: contig of 1452 bp in length  
\* 129052 129151: gap of unknown length  
\* 129152 130882: contig of 1731 bp in length  
\* 130883 130983: gap of unknown length  
\* 130984 132045: contig of 1063 bp in length  
\* 132046 132145: gap of unknown length  
\* 132146 133485: contig of 1340 bp in length  
\* 133486 133585: gap of unknown length  
\* 133586 135588: contig of 2003 bp in length  
\* 135589 135689: gap of unknown length  
\* 135690 137050: contig of 1362 bp in length  
\* 137051 137150: gap of unknown length  
\* 137151 138633: contig of 1483 bp in length  
\* 138634 138733: gap of unknown length  
\* 138734 140308: contig of 1575 bp in length  
\* 140309 140408: gap of unknown length  
\* 140409 141706: contig of 1298 bp in length  
\* 141707 141806: gap of unknown length  
\* 141807 143314: contig of 1508 bp in length  
\* 143315 143414: gap of unknown length  
\* 143415 144796: contig of 1382 bp in length  
\* 144797 144896: gap of unknown length  
\* 144897 146087: contig of 1191 bp in length  
\* 146088 146187: gap of unknown length  
\* 146188 147644: contig of 1457 bp in length  
\* 147645 147744: gap of unknown length  
\* 147745 149005: contig of 1261 bp in length  
\* 149006 149105: gap of unknown length  
\* 149106 150998: contig of 1893 bp in length  
\* 150999 151098: gap of unknown length  
\* 151099 152445: contig of 1347 bp in length  
\* 152446 152545: gap of unknown length

Query Match 20.7%: Score 228.2; DB 2: Length 167150;  
Best Local Similarity 87.9%: Pred. No. 6.8e-36;  
Matches 284; Conservative 0; Mismatches 33; Indels 6; Gaps 3;

QY 724 atctgtgaagcatalcccccagctccctccatccacagaatgtctactctgtgtctccacag 783  
Db 59881 ATGCGTCTTCTTAACCCCAAGTCCCTTCTACAGTGGTACTACGGGCTTCCACAG 59940

QY 784 gagaagactccatctcattatccacccctcttgaaggacat----acttgcct 839  
Db 59941 GA-AAGGCTCCCATCTTACCTTATTCACCCCTTTGGAAAGCAATTCATTTGCT 59999

QY 840 tttcaagaaggttgagctgttcaagttcacaagggaagaagatcctctccctc 899  
Db 60000 TGTCAAGAGGTGGCTTGTTCAGGTCTACAGGGGAAGGCTGAAATCTCTCCCTC 60059

QY 900 atgggtgtgttgagccttaaccggaataatgtcct-ggctaggagggagtgaagaca 958  
Db 60060 ATGATGTTTATAGGCCCTTAATGGAAGAAATGTCTCGGCTGAGGAGGAGGAGACC 60119

QY 959 tggccagctatccttaagcccaagaacacaaatgtctcacaacaccacataaaggact 1018  
Db 60120 TGGCCACACTGTGCTTACTCTCGAAGAACCAAAATGTCTCAAACTACCAATGAAGACT 60179

QY 1019 ctcctgttgagcaccagagaaa 1041  
Db 60180 CCGCTTGGTAGCACACAGAGAA 60202

RESULT 13  
AK024095



LOCUS AK024095 1841 bp mRNA linear PRI 29-SEP-2000  
 DEFINITION Homo sapiens cDNA FLJ14033 fis, clone HEMBA1004479, weakly similar  
 to HPOX1A-INDUCIBLE FACTOR 1 ALPHA.  
 ACCESSION AK024095  
 VERSION AK024095.1 GI:10436389  
 KEYWORDS oligo capping, fis (full insert sequence).  
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to  
 mRNA, clone\_id:HEMBA1 clone:HEMBA1004479.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Shiraatori,A., Sudo,H.,  
 Nagatsuna,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
 Nakamura,Y., Nagahari,K., Masuko,Y., Nishimura,K. and Iwayanagi,T.  
 NEDO human cDNA sequencing project  
 TITLE Unpublished (2000)  
 JOURNAL 2 (bases 1 to 1841)  
 REFERENCE Isogai,T. and Otsuki,T.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1841  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEMBA1004479"  
 /tissue\_type="whole embryo, mainly head"  
 /clone\_lib="HEMBA1"  
 /dev\_stage="embryo, 10 weeks"  
 /note="cloning vector: pME18SFL3"  
 119..832  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB14824.1"  
 /db\_xref="GI:10436390"  
 /translation="MRPAGAAARRPCTSWLTRCPSPAPASAPFTWRPLSCASPSATC  
 ACTASAPQELDALPQOTLSRRKVEAPTERFSLRMKSTLRSRGTVLTKATWV  
 LNCSGHRAVKKPAPSPAPSPDSEPLCLVLCALPHPGSLTSEFVNLREGDMA  
 YLSEVNSHIGLSOLELIGHSTFDFIHPDQDELQDALPLHLNTSSLLKPGQTVSEFL  
 APSYVPRFSRPHLPWPMP"  
 CDS  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB14824.1"  
 /db\_xref="GI:10436390"  
 /translation="MRPAGAAARRPCTSWLTRCPSPAPASAPFTWRPLSCASPSATC  
 ACTASAPQELDALPQOTLSRRKVEAPTERFSLRMKSTLRSRGTVLTKATWV  
 LNCSGHRAVKKPAPSPAPSPDSEPLCLVLCALPHPGSLTSEFVNLREGDMA  
 YLSEVNSHIGLSOLELIGHSTFDFIHPDQDELQDALPLHLNTSSLLKPGQTVSEFL  
 APSYVPRFSRPHLPWPMP"  
 BASE COUNT 404 a 514 c 519 g 404 t  
 ORIGIN  
 Query Match 19.2%; Score 211.6; DB 9; Length 1841;  
 Best local Similarity 83.1%; Pred. No. 7.5e-33;  
 Matches 241; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 472 caaaggaactcaagcgcctgaccccaagcgcgaacctgtcaagaagaagctgaa 531  
 ||||||| ||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 272 CAAGAGAGACTTCAGACGCCCTGACCCACGACACCTGTGCCAGAGAGAGGTGAG 331  
 QY 532 gccccaacagagcgcaactttccctcggaatgaagaagcagctaccagcagaaggcgc 591  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 332 GCCCACCAGGAGGGTCTCTCTGCGCATGAAGAGTACACACCCACCGCGCGC 391  
 QY 592 agcgtcaacctaaagcgacacctgaaggtgctgacatcagacacatagaagggc 651  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 392 ACCCTCAACTCAAGCGGCCACCTGGAAGGTGCTGAACATGCTGTGACATATGAGGCC 451

QY 652 tacaagccctgacagacttccctccgagagccctcgtccgacccctcga 711  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 452 TACAGCCACTGCGACAGACTTTCACAGCTGGAGGCCCTGACAGACCCCGCTCAG 511  
 QY 712 tgcctggtctattctgtgaagcattcccccagctcccttcacagatgg 761  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 512 TGCCTGCTGCTCACTCGAAGACCATCCACCCAGCAGACCTGAGGG 561  
 RESULT 14  
 AC079550/c 134189 bp DNA linear HTG 02-SEP-2000  
 LOCUS AC079550/c  
 DEFINITION Mus musculus clone RP23-422A19, WORKING DRAFT SEQUENCE, 30  
 unordered pieces.  
 ACCESSION AC079550  
 VERSION AC079550.1 GI:9964915  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DNAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 134189)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Mouse  
 JOURNAL Unpublished  
 2 (bases 1 to 134189)  
 REFERENCE DOE Joint Genome Institute.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 1898939  
 Center clone name: RPCI-23\_422A19  
 -----  
 Summary Statistics  
 Consensus quality: 105629 bases at least Q40  
 Consensus quality: 114783 bases at least Q30  
 Consensus quality: 117545 bases at least Q20  
 Estimated insert size: 208000; agarose-fp estimation  
 Estimated insert size: 131289; sum-of-confids estimation  
 Quality coverage: 8.71 in Q20 bases; agarose-fp estimation  
 Quality coverage: 13.79 in Q20 bases; sum-of-confids estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 1340 1339: contig of 1339 bp in length  
 \* 1439: gap of unknown length  
 \* 1440 2741: contig of 1302 bp in length  
 \* 2742 2841: gap of unknown length  
 \* 2842 4408: contig of 1567 bp in length  
 \* 4409 4508: gap of unknown length  
 \* 4509 5699: contig of 1191 bp in length  
 \* 5700 5799: gap of unknown length  
 \* 5800 7306: contig of 1507 bp in length  
 \* 7307 7406: gap of unknown length  
 \* 7407 8863: contig of 1457 bp in length  
 \* 8864 8963: gap of unknown length  
 \* 8964 10169: contig of 1206 bp in length  
 \* 10170 10269: gap of unknown length  
 \* 10270 11579: contig of 1310 bp in length  
 \* 11580 11679: gap of unknown length  
 \* 11680 12992: contig of 1313 bp in length  
 \* 12993 13092: gap of unknown length



Query Match	Best Local Similarity	17.8% Score 196: DB 2: Length 134189;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
Db 130269	AGGTGCAACACCGAGCTGCTTACAGCTGGCGCACACTGTCCTTTGGCGCGGCGTACAC	130210
Db 130209	CAGGAGACGGAGGTCTTACAGCTGGCGCACACTGTCCTTTGGCGCGGCGTACAC	130150
Db 130149	GGCGACCTGGACAAAGGCTCCATCTATGCGCCCTCAATCAGTACTGCGATGACCGC	130090
Db 130089	CTCTGCGCAGCAGTGG	130074

AF304864	LOCUS	AF304864	3605 bp	mRNA	linear	VRT 05-JUN-2001
	DEFINITION	Oncorhynchus mykiss hypoxia-inducible factor 1 alpha (HIF-1a) mRNA.				
	ACCESSION	AF304864				
	VERSION	AF304864.1	GI:13561505			
	KEYWORDS					
	SOURCE	rainbow trout.				
	ORGANISM	Oncorhynchus mykiss				
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.				
	AUTHORS	1 (bases 1 to 3605) Soltano,A.J., Rabergh,C.M.I., Gassmann,M., Sistonen,I. and Nikimaa,M.				
	TITLE	Characterization of a Hypoxia-Inducible Factor (HIF-1 alpha) from Rainbow Trout: Accumulation of Protein Occurs at Normal Venous Oxygen Tension				
	JOURNAL	J. Biol. Chem. 276 (23), 19699-19705 (2001)				
	PUBMED	11278461				
	RESEARCH	2 (bases 1 to 3605)				
	AUTHORS	Soltano,A.J., Rabergh,C.M.I., Gassmann,M., Sistonen,I. and Nikimaa,M.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (12-SEP-2000) Department of Biology, Laboratory of Animal Physiology, University of Turku, Yliopistomaki, Turku FIN-20014, Finland				
	FEATURES					
	source	Location/Qualifiers				
	gene	1..3605				
		/organism="Oncorhynchus mykiss"				
		/db_xref="taxon:8022"				
		1..3605				
		/gene="HIF-1a"				
		301..2601				
		/gene="HIF-1a"				
		/note="bHLH/PAS A/B transcription factor"				
		/codon_start=1				
		/product="hypoxia-inducible factor 1 alpha"				
		/protein_id="AAK30364.1"				
		/db_xref="GI:13561506"				
		/translation="MDTGVPEPKSRVSSDRRREKSRDARCRKRESEVFELAOEILPLRSVNTLDKASIMRAIISTYLMRNILSDNBEQGERENDSLNSYLAIGFVLSVDGMITLSENVKCLGIAQIDLGLSVFETHPCDHELEEMLVHRTGSKKKEPNTRESEFPLRMKCTLNNRGTNRVKSATWVKSATVHVESPAIDPGHREPVPYLVLCDDPIPHPSNIENAPLDTKTFLSRHITLDMKFTYCDERTELMGVDEPDILNRVVEYHALDSHLMKTHNLPAKGVGSGOYRLMKRGCFVWETQATYVYNNKSNOCVQCVNVYVSGIEBEKKMLSLQEDNEDRPVKKLEEESSPEPVSLIKEEKSEIDYIKLTRAVETQPLSLYDRKEPEPALVTLIAPAGTITISLSSPSDSITLQKEVILYKVMLPSTSDKTLALPLSLPPSDOHLVPNTSVPTTSTGSDSSVPGSHSTPEEPISPLDFCFPEMESDINAEFKLIDWETLFAINPEKPFPTLQAMEDLDLEMLAPIPMDDIFOLRLISPEEPLSCGPAQPLECCSLSCSVRLQEVHVSVPFNAGSLTASIPAAADIPALAPADSPRCASALITKTYVPMQDREISLRSLSAONAORKKRSLSOAVIGIGLLDHPRGKRLKXVSELSEHADAPNRTITLPTDLASNLGLISSGSSPPTLPOLUTRYDQEVNAPVGGRIQLLQGEELLSALDOVN"				
	BASE COUNT	888 a 951 c 866 g 900 t				
	ORIGIN					
	Query Match	17.5%; Score 192; DB 5; Length 3605;				
	Best Local Similarity	59.1%; Pred. No.7.8e-29;				
	Matches 436; Conservative	0; Mismatches 220; Indels 82; Gaps				
0Y	43	aggtgtaacaccggaactctcggaagaagaatgtcggggagacggccgcagcgcgccagc	102			
Db	334	AGGGTGAAGCTCGGACCGGAGGAAGAGAGTCCGGGAGTGGCGCGAGAGCAGAGAGGGG	393			
0Y	103	cagggaagcgaaggtctgtaaccagctggcgacacatctcccttggcgcggtcagc	162			
Db	394	AAGGAATCTGAAGGTCTTCTACGAGCTGGCCCGAGAGCTCCCTCGCCACACAGCTCAC	453			
0Y	163	gcgcacctggacaaggtccatcatctgcgctcaacaatcagctactctgcgcatg-----	216			
Db	454	TCCAACTGTGGAAGAGGCTCCATCTGAGAGCTGCGCATCACTTACTTGCATAGAGGAAT	513			

```

QY 217 -----cacgcctctgcgaagcagtgtaaaagggagagagccactggac 262
Db 514 CTGCTGACGACACACATATAGAGGAGGAGGAGGAGGAAATGACCTCCAGCTGAAT 573
QY 263 gacctgactaagagccctggaaggttcgtcatgtactcacgcgcgagggagacatg 322
Db 574 GGATCCTATCTGAAGGCTTATAGAGGCTTCCTCATGTGCTGCTCTGAGGACGGGACATG 633
QY 323 gcttacctgtcgaaatgtcagcaagcctgggcctcagtcagtgaaacctctgtcct 382
Db 634 ATTTATCTCTGAGAAATGTCAACAAGTGTGGGTCTGGCACAG----- 678
QY 383 cctccctgatacaaacccactcctgtaccaatctctctcgtgagctcatgtgacaca 442
Db 679 -----ATTGACTGACAGACTTA 697
QY 443 gtaacttgattatcaccctcgtgacaaaggaacttcaagaagccttgacccca 502
Db 698 GTGTGTTGAGTACACACACCCCTGTGACCATGAGAGCTGAGAGAGATGCTGTACACA 757
QY 503 ggcgcgaacctgtcaagaagaagctggaagccccaacagaagcgccactttccctcgaa 562
Db 758 GAACAGGACCTTAATAAAGTCCAAAGAACCAACACAGACGCAAGCTTCTTTCGGGA 817
QY 563 tgaagagcagcgttcaaccagcagagcgcgacgttcaacctcaagcgcgcaacctggaag 622
Db 818 TGAATGTACCTCCTACCAACAGGAGGCGGACACTGTCAATGTCAATACAGCACCTGGAAG 877
QY 623 tgcgtgactgctcagacatalgaaggccttacaagcccccgtgcacagacttccctgcg 682
Db 878 TTCTCCACTGCTCAGACATGTGCGGCTCATGAATCCCAGCTGAGAGATCCCGGTG 937
QY 683 ggaagccctgctccgagcctccctgcaatgctgtgttatactgtgaagcattccccc 742
Db 938 GG-----CACAGGAGCGCGTCCCTACCTGCTGTGTGTGACCCCATCCCCC 991
QY 743 agtcccttcaagatg 760
Db 992 ACCCTCCAAATCGAGG 1009

```

Search completed: September 8, 2002, 07:29:17  
 Job time: 6900 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 05:58:42 ; Search time 49.21 Seconds  
(without alignments)  
692.942 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALLGLRVSNTFLRKESR.....TESSLPSTWVLMALNRKRCG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855.5	52.3	662	20	AAV06295
2	831	50.8	632	22	AAB93326
3	573	35.0	648	22	AAB94934
4	518.5	31.7	199	22	AAB93710
5	495	30.3	245	21	AAV94627
6	495	30.3	330	21	AAV94628
7	495	30.3	642	21	AAV84168
8	495	30.3	652	21	AAV94629
9	495	30.3	669	21	AAV84167
10	495	30.3	697	21	AAV84166
11	495	30.3	701	21	AAV84173

12	495	30.3	710	21	AAV84172
13	495	30.3	724	21	AAV84171
14	495	30.3	749	21	AAV84170
15	495	30.3	789	21	AAV84169
16	495	30.3	813	21	AAV94636
17	495	30.3	826	18	AAV06557
18	495	30.3	826	19	AAV80418
19	495	30.3	826	20	AAV06289
20	495	30.3	826	20	AAV94640
21	495	30.3	826	21	AAV69407
22	495	30.3	826	22	AAV6854
23	477.5	29.2	875	19	AAV37098
24	467	28.5	870	20	AAV06290
25	466	28.5	485	21	AAV79160
26	466	28.5	870	19	AAV37097
27	466	28.5	870	21	AAV79161
28	444	27.1	373	18	AAV06559
29	444	27.1	805	18	AAV06558
30	426	26.0	237	22	AAV95649
31	345.5	21.1	613	21	AAV94630
32	323	19.7	373	22	ABG20871
33	323	19.7	756	21	AAV94635
34	294	18.0	590	19	AAV68091
35	293.5	17.9	594	19	AAV68092
36	281	17.2	1507	22	ABV58562
37	278.5	17.0	691	20	AAV06294
38	269	16.4	776	22	ABG28236
39	265	16.2	634	22	ABV71624
40	265	16.2	657	18	AAV30758
41	264	16.1	828	22	ABG20868
42	251	15.3	958	22	ABV62764
43	250.5	15.3	924	21	AAV52991
44	222.5	13.6	99	21	AAV54411
45	214	13.1	74	21	AAV94626

#### ALIGNMENTS

RESULT 1	
AAV06295	standard; Protein: 662 AA.
AAV06295:	
AC	23-AUG-1999 (first entry)
DT	Mouse transcription regulator MOP7.
XX	
DE	MOP7; member of the PAS superfamily; bHLH-PAS; mouse; transcription regulator; hypoxia inducible factor 3 alpha.
KW	
KX	
OS	Mus musculus.
XX	
PN	W09928464-A2.
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-US25314.
XX	
PR	28-NOV-1997; 97US-0066863.
XX	
PA	(WISC) WISCONSIN ALUMNI RES FOUND.
XX	
PI	Bradfield CA, Gu YZ, Hogenesch JB;
XX	
DR	WPI: 1999-371120/31.
DR	N-PSDB: AAX58986.
XX	
PT	Developmental signal transduction associated proteins
XX	
PS	Claim 6; Page 101; 106pp; English.
XX	

CC The present sequence represents mouse MOP7, a novel member of the  
CC PAS superfamily, where PAS stands for PER/ARNT/SIM domains. MOP7  
CC cDNA (see AAX58986) was identified in a search of murine ESTs designed  
CC to identify bHLH-PAS proteins, and by RACE amplification of lung  
CC cDNA. MOP7 was characterised as hypoxia inducible factor 3 alpha  
CC (HIF 3 alpha). Its expression profile is distinct from that of  
CC HIF 1 alpha (see AY06289), HIF 2 alpha (see AA006290), MOP3 (see  
CC AY06291), Ah receptor and Ah receptor nuclear translocator (ARNT),  
CC suggesting a different functional role. MOP7 probably regulates  
CC the same genes as HIF 1 alpha and 2 alpha, as evidenced by its  
CC dimerisation with the same partners (ARNT, MOP3) and recognition  
CC of the same core response element. MOP7 may have a functional  
CC role associated with response to low oxygen in the tissues in  
CC which it is expressed. The invention provides novel MOPs 2-9  
CC nucleic acids (see AAX58981-88) and proteins (see AY06289-97).  
CC These are useful in a variety of research, diagnostic and  
CC therapeutic applications. Several of the MOPs are alpha-class  
CC hypoxia-inducible factors. Others are involved in circadian signal  
CC transduction.

XX Sequence 662 AA;

Query Match 52.3%; Score 855.5; DB 20; Length 662;  
Best Local Similarity 68.4%; Pred. No. 7.6e-80; Indels 55; Gaps 4;  
Matches 182; Conservative 5; Mismatches 24;

YY 9 RSMTELKRSRDAARSRQETEVLYQLAHTLPFARGVSAHLDKASIMRLTSLRMHR 68  
DB 7 rsntelkrsrdaarrrrsgelevlyqlahltlpfargvsahldkasimrltstlrmhr 66  
YY 69 LCAAGKRGATGRLLPBPGGFRHGRHGRHGLPVGKCOQARPGOSVDIC----- 120  
DB 67 lcaagewn-----qvekgyepldacylkalegf 94  
YY 121 -----SSSLIHNPFGTNGFS-LELIGHSITPDTFHPCDDELQDALPRMLSKKL 170  
DB 95 vmyltaegdmaylsevenykhlglsqlelghnsifdihpcdeeqdalprmlskkl 154  
YY 171 EAAPERHFSLRMKSTLTSRGRTLNKAAATWKVLHSGHMRAYKPPAOTSPAGSPSEPPL 230  
DB 155 eaaperhfslrmskstltsrgrtlntkaatwkvlnhsgmhraykppatqspagspseppl 214  
YY 231 QCLVLCGAIP-----QLPFHDGATL 251  
DB 215 qclvllicealpaspelprlggrgafll 240

RESULT 2

AAB93326 standard; Protein: 632 AA.

XX AAB93326;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12422.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
PS Claim 8; SEQ ID 12422; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-PT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.

XX Sequence 632 AA;

Query Match 50.8%; Score 831; DB 22; Length 632;  
Best Local Similarity 68.2%; Pred. No. 2.5e-77;  
Matches 178; Conservative 11; Mismatches 30; Indels 42; Gaps 5;

YY 1 MALGLQVRNSNTELRKRSRDAARSRQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
DB 1 malglqrarsttelrkrksrdaarrrrsgelevlyqlahltlpfargvsahldkasimrlt 60

YY 61 ISYLRMHRLLCAAGKRGATGRLLPBPGGFRHGRHGRHGLPVGKCO----- 109  
DB 61 isylrmhrllcaag-----ewngvgaggepldacylkalegfymvl 100

YY 110 QAPGQSVDCSSSLIHNPFGTNGFS-LELIGHSITPDTFHPCDDELQDALPRMLSKKL 168  
DB 101 taegdmay-----lsevenykhlglsqlelghnsifdihpcdeeqdalprmlskkl 154

YY 169 KLEAPERHFSLRMKSTLTSRGRTLNKAAATWKVLHSGHMRAYKPPAOTSPAGSPSEP 228  
DB 155 kveaperhfslrmskstltsrgrtlntkaatwkvlnhsgmhraykppatqspagspsep 214

YY 229 PLQCLVLCGAIPQOLPFHDGATL 249  
DB 215 plqclvllicealp-----hpgs 231

RESULT 3

AAB94934 standard; Protein: 648 AA.

XX AAB94934;

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX 26-JUN-2001 (first entry)

XX		Human protein sequence SEQ ID NO:16415.
DE		
XX		Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW		
OS		Homo sapiens.
XX		
PX		Epi074617-A2.
PD		
XX		07-FEB-2001.
PF		
PE		28-JUL-2000; 2000EP-0116126.
XX		
XX		29-JUL-1999; 99JP-0248036.
PR		27-AUG-1999; 99JP-0300253.
PR		11-JAN-2000; 2000JP-0118776.
PR		02-MAY-2000; 2000JP-0183767.
PR		09-JUN-2000; 2000JP-0241899.
PA		(HELI-) HELIX RES INST.
PI		Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI		Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR		WPI: 2001-318749/34.
PT		Primer sets for synthesizing polynucleotides, particularly the 5602
PT		full-length cDNAs defined in the specification, and for the detection
PT		and/or diagnosis of the abnormality of the proteins encoded by the
PT		full-length cDNAs -
XX		
PS		Claim 8; SEQ ID 16415; 2537pp + CD ROM; English.
XX		
CC		The present invention describes primer sets for synthesizing 5602
CC		full-length cDNAs defined in the specification. Where a primer set
CC		comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC		to the complementary strand of a polynucleotide which comprises one of
CC		the 5602 nucleotide sequences defined in the specification, where the
CC		oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC		of an oligonucleotide comprising a sequence complementary to the
CC		complementary strand of a polynucleotide which comprises a 5'-end
CC		sequence and an oligonucleotide comprising a sequence complementary to a
CC		polynucleotide which comprises a 3'-end sequence, where the
CC		oligonucleotide comprises at least 15 nucleotides and the combination of
CC		the 5'-end sequence/3'-end sequence is selected from those defined in
CC		the specification. The primer sets can be used in antisense therapy and
CC		in gene therapy. The primers are useful for synthesizing polynucleotides,
CC		particularly full-length cDNAs. The primers are also useful for the
CC		detection and/or diagnosis of the abnormality of the proteins encoded by
CC		the full-length cDNAs. The primers allow obtaining of the full-length
CC		cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC		AAH13653 to AAH18742 represent human cDNA sequences; AAB92446 to
CC		AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC		represent oligonucleotides, all of which are used in the exemplification
CC		of the present invention.
XX		
SO		
Sequence	648 AA:	
Query Match	35.0%;	Score 573; DB 22; Length 648;
Best Local Similarity	61.0%;	Fred. No. 1.8e-50;
Matches 125;	Conservative 11;	Mismatches 27; Indels 42; Gaps 5;
QY	57	MRLTISYLEMHLCAAGCGKRGATGTLPEGGGFGFHRRGRGLPGKCO----- 109
DB	1	MLITLSYIMHILCAAG-----ewngvgaagepdiacYkaIegf 40
QY	110	----QAPGPOSVDLCSLSIHNPFTGTNFS-LELIGHSIFDIHPDDELALPRRN 164
DB	41	vmlvalsgdmay-----Isenvskhlgsqlelghsisfdiflnpodgeidatlpqgt 94
QY	165	ISKKLLEAFTEHFESLRMKSTLTSGRTINTLKAAATWKVLHGSGHMRAKYPPAQTSPAGSP 224

Db 95 Isrlkveaptercrlslmkslersgrtllnlhkaatw\*VIncsjhmrayRppaqtsppasp 154

Oy 225 RSEPPQLCTVLICELAIPLQLPDPHDA 249  
|||||  
155 dsepplgclvllcealp-----hpgs 175

RESULT 4  
AAB93710  
ID AAB93710 standard; Protein: 199 AA.  
XX AAB93710:  
AC  
XX 26-JUN-2001 (first entry)  
DT  
XX Human protein sequence SEQ ID NO:13303.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW  
XX Homo sapiens.  
OS  
XX EPI074617-A2.  
PN  
XX EPI074617-A2.  
PD  
XX 07-FEB-2001.  
PF  
XX 28-JUL-2000; 2000EP-0116126.  
PR  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
PA (HELI-) HELIX RES INST.  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
DR WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length CDNs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length CDNs -  
XX  
XX Claim 8: SEQ ID 13303; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length CDNs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides  
CC particularly full-length CDNs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length CDNs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13632 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX  
XX Sequence 199 AA;  
SQ

	Query Match	31.7%	Score 518.5	DB 22	Length 199
	Best Local Similarity	68.0%	Pred. No. 1.5e-45		
	Matches 104	Conservative 9	Mismatches 13	Indels 27	Gaps 2
QY	120	CSSSLHNPQGTNF-----	SLFLGHSIRPFTHPCQOELQ	156	
		:    : :  :		73	
Db	14	ctslvrlpnpaspaasaprtwtprlscaspsatcactaapqjlelignstfdlfnpcqeeiq			
QY	157	DALVPRPRLNSKKKLAEPTERRHFSLSRMSSTLTSRGRLNLKAATWVYLHSCGMRAVKKPPA	216		
		: : :  : : :  : : :  : : :  : : :  : : :  : : :  : : :  : : :			
Db	74	dalppgqlarlkveaprtcrlfslmksltlstrgrlrlnkattwkvlnscsglmraykppa	133		
QY	217	QTPAGSPRSEPPQLQVLICELAIPOLEPHDGA	249		
Db	134	qtspagspdssepplqclvlceaiP-----hpqs	162		
RESULT, 5					
QY	157	AAAY94627	standard; protein; 245 AA.		
XX	AC	AAAY94627;			
XX	DT	15-AUG-2000	(first entry)		
XX	DE	HIF-1alpha variant protein sequence HIF-1alpha/1-245.			
KW	XX	Hypoxia-inducible factor 1alpha: HIF-1alpha; PAS-B; N-TAD; C-TAD;			
XX	XX	regulation; angiogenesis; erythropoiesis; glycolysis; human.			
OS	XX	Homo sapiens.			
PN	XX	MO200029437-A1.			
PD	XX	25-MAY-2000.			
XX	XX	11-NOV-1999;	99MO-SE02053.		
PR	XX	13-NOV-1998;	98SE-0003891.		
PA	XX	(PHAA ) PHARMACIA & UPJOHN AB.			
XX	XX	Berkenstam A, Poellinger L;			
DR	XX	WPI; 2000-399715/34.			
PT	XX	Human hypoxia-inducible factor alpha variants for identifying compounds			
PT	XX	that modulate its functional domain and regulate genes involved in			
PT	XX	angiogenesis, erythropoiesis -			
PS	XX	Claim 13: Page 68-69; 87pp; English.			
XX	XX	This sequence represents a fragment of the hypoxia-inducible factor			
CC	XX	(HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha			
CC	XX	is a multi-step process which includes hypoxia-dependent nuclear import			
CC	XX	and activation of the transactivation domain. The HIF-1alpha consists of			
CC	XX	a number of functional domains including a PAS-B (Per. Ant. Sim) domain			
CC	XX	located in human HIF-1alpha between amino acids 173 and 390, a C-terminal			
CC	XX	nuclear localization sequence located at amino acids 718-584, a			
CC	XX	transactivator domain (N-TAD) located between amino acids 531 and 584,			
CC	XX	and a second transactivator domain (C-TAD) located between 813 and 826.			
CC	XX	The invention relates to isolated variants of HIF-1alpha, such as that			
CC	XX	represented by the present sequence. The variants are useful for			
CC	XX	identifying compounds capable of modulating the function of a functional			
CC	XX	domain of human HIF-1alpha. The method comprises contacting a candidate			
CC	XX	compound with a cell expressing a HIF-1alpha variant conjugated to a			
CC	XX	molecular probe. The localization of the probe can be detected in the			
CC	XX	cell. The Aequorea victoria green fluorescent protein can be used as the			
CC	XX	molecular probe. The compounds are useful for the regulation of			
CC	XX	HIF-1alpha target genes, such as those involved in the regulation of			
CC	XX	angiogenesis, erythropoiesis and glycolysis.			
XX	XX	Sequence 245 AA:			

Query Match	30.3%	Score 495	DB 21	Length 245
Best Local Similarity	46.2%	Pred. No. 5,6e-43		
Matches 114	Conservative 26	Mismatches 59	Indels 48	Gaps
QY	11	NIETLREKSRDARARRSQETEVVLQATLTPRPARVSANHLKASIMRLTISLYBMHRLC	70	
DB	14	sserlkeksrdaararrrsqetevvlyeladqlrphnvsyhlkkaasymrltlislylvrxll	73	
QY	71	AAGKRGKRGATRGRLDEGPGFRRGTHRRGRHGLPYVKSCOA-----DQPOSVDLC	120	
DB	74	deg-----dlldiedamkagmncfykaldgfvwmvld	105	
QY	121	SSSLIH-----NPTGQ-TNFSLELIGHSTFDTIHPQDQDELODALTLPRLNSKKKLEAET	174	
DB	106	dqdmlyisdvnhkynqlqf--elghsvfdftlhpdcbeemrcemilthrgilvkkqkqnt	163	
QY	175	EHFHSIRMKSTLTSGKRTNLKATQTMVYHCGSHGMAYKRPRAQTSPPAGSPRSEPPLOCIV	234	
DB	164	qsfllmkcrltsrgtmmiksalwkvlnhcgahlvnydtnsqpgcg--ykkpmtctiv	221	
QY	235	LICEAIP 241		
DB	222	lliceip 228		
RESULT 6				
AAV94628				
ID	AAV94628	standard; protein; 330 AA.		
XX	AAV94628;			
AC				
XX				
DT	15-AUG-2000	(first entry)		
DE				
XX	HIF-1alpha variant protein sequence HIF-1alpha/1-330.			
XX				
KM	Hypoxia-inducible factor 1alpha; HIF-1alpha; PAS-B; N-TAD; C-TAD;			
KW	regulation; angiogenesis; erythropoiesis; glycolysis; human.			
OS	Homo sapiens.			
PN	WO200029437-A1.			
PD	25-MAY-2000.			
XX				
PF	11-NOV-1999; 99WO-SE02053.			
XX				
PR	13-NOV-1998; 98SE-0003891.			
PA	(PHAA ) PHARMACIA & UPJOHN AB.			
XX				
PI	Berkenstam A, Poellinger L;			
XX				
DR	WPI: 2000-399715/34.			
XX				
PT	Human hypoxia-inducible factor alpha variants for identifying compounds			
PR	that modulate its functional domain and regulate genes involved in			
XX	angiogenesis, erythropoiesis -			
XX				
PS	Claim 15; Page 69-70; 87pp; English.			
XX				
CC	This sequence represents a fragment of the hypoxia-inducible factor			
CC	(HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha			
CC	is a multi-step process which includes hypoxia-dependent nuclear import			
CC	and activation of the transactivation domain. The HIF-1alpha consists of			
CC	a number of functional domains including a PAS-B (Per, Arnt, Sim) domain			
CC	located in human HIF-1alpha between amino acids 173 and 390, a C-terminal			
CC	nuclear localization sequence located at amino acids 718-584, a			
CC	transactivator domain (N-TAD) located between amino acids 531 and 584,			
CC	and a second transactivator domain (C-TAD) located between 813 and 826.			
CC	The invention relates to isolated variants of HIF-1alpha, such as that			
CC	represented by the present sequence. The variants are useful for			

XX Creating hypoxia of ischemia in  
PS Claim 1; Page -; 96pp; English

PA 13 NOV 1950, 303E 0003031.  
XX  
XX (PHAA ) PHARMACIA & UPJOHN AB  
XX  
PI Berkenstam A. Poellinger I:

XX WPI: 2000-399715/34.  
DR  
XX  
XX Human hypoxia-inducible factor alpha variants for identifying compounds  
PT that modulate its functional domain and regulate genes involved in  
XX anglogenesis, erythropoiesis .  
XX  
XX Claim 15; Page 69-70; 87pp; English.  
XX  
XX This sequence represents a fragment of the hypoxia-inducible factor  
CC (HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha  
CC is a multi-step process which includes hypoxia-dependent nuclear import  
CC and activation of the transactivation domain. The HIF-1alpha consists of  
CC a number of functional domains including a PAS-B (Per, Arnt, Sim) domain  
CC located in human HIF-1alpha between amino acids 173 and 390, a C-terminal  
CC nuclear localization sequence located at amino acids 718-584, a  
CC transactivator domain (N-TAD) located between amino acids 531 and 584,  
CC and a second transactivator domain (C-TAD) located between 813 and 826.  
CC The invention relates to isolated variants of HIF-1alpha, such as that  
CC represented by the present sequence. The variants are useful for  
CC identifying compounds capable of modulating the function of a functional  
CC domain of human HIF-1alpha. The method comprises contacting a candidate  
CC compound with a cell expressing a HIF-1alpha variant conjugated to a  
CC molecular probe. The localization of the probe can be detected in the  
CC cell. The Aequorea victoria green fluorescent protein can be used as the  
CC molecular probe. The compounds are useful for the regulation of  
CC HIF-1alpha target genes, such as those involved in the regulation of  
CC anglogenesis, erythropoiesis an glycolysis.  
XX  
SQ Sequence 652 AA:  
  
Query Match 30.3%; Score 495; DB 21; Length 652;  
Best Local Similarity 46.2%; Pred. No. 2,3e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;  
  
OY 11 NPELREKSRDAARSRQSETEVLYQLAHTLPFAGVSAHLDKASTIMRLTISYLMNRHC 70  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 14 sseRTkEksrdaArstrsEseVfYelaNqDlRlphnvshldkaSwmrltIsyIvvrKl1 73  
OY 71 AAGGKRGATGRILDEPGGFRNGTHRRGRHGLPYGKCOQA-----PGQSVYDLC 120  
DB 74 dAg-----dlIdedmkkaqmcfyIkaldgfymvltD 105  
OY 121 SSSLIH-----NPPRG-TNFSLELIGHSTFDFTINPCDOEELDALTPRNLSSKKLEAFT 174  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 106 dgdmiYisdvNkYmgltqf--elcgshvfdftphodheememlthnglvkkgqeqt 163  
OY 175 EHNFSLRMKSSTLTSGRNTLNKAATWVKVYHSCGHNMRAYKRPRAOTSPAGSPRSEPRLOCTV 234  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 164 gsfIfImkcItIsrgTmnlksatwkvhNctghlhvYdtSnqpCg--yKkprmtctIv 221  
OY 235 LTCEAIP 241  
DB :::::::::::::::::::::  
DB 222 lIcepIp 228  
  
RESULT 9  
AAV84167  
ID AAV84167 standard: Protein; 669 AA.  
XX  
XX AAV84167;  
XX  
XX 03-JUL-2000 (first entry)  
XX  
XX A variant of human hypoxia inducible factor-1 alpha protein.  
XX  
XX Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
XX hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
XX ischemia related damage; anglogenesis; coronary artery disease;  
XX ischemic tissue damage.  
XX  
OS Synthetic.

OS Homo sapiens.  
XX  
XX WO200010578-A1.  
XX  
XX 02-MAR-2000.  
XX  
XX 25-AUG-1999; 99WO-US19416.  
XX  
XX 25-AUG-1998; 98US-0148547.  
XX  
XX (UYTO ) UNITV JOHNS HOPKINS SCHOOL MEDICINE.  
XX  
XX Semenza GL;  
XX  
XX WPI: 2000-246493/21.  
XX  
XX Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for  
PT treating hypoxia or ischemia-related tissue damage .  
XX  
XX  
XX Claim 1; Page -: 96pp; English.  
XX  
XX The present sequence represents a variant of hypoxia-inducible factor  
CC (HIF)-1 alpha, comprising amino acids 1-391 and 549-826 of the wild  
CC type protein (see AAV69407). The HIF-1alpha variants are stable under  
CC hypoxic and non-hypoxic conditions. The variants comprises amino acid  
CC residues 1-391 and 521-826, 549-826, 576-826, 429-826, 494-826,  
CC 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha  
CC polypeptide, in which residues 551 and 552 are not serine and threonine,  
CC respectively. The HIF-1alpha variant polynucleotide sequences are useful  
CC for increasing expression of a hypoxia inducible gene in a cell. They  
CC is also useful for providing constitutive expression of a hypoxia  
CC inducible factor in a cell, and for reducing or preventing hypoxia or  
CC ischemia related damage. The variant HIF-1alpha polypeptides are useful  
CC for providing prophylactic therapy for inducing the level of  
CC anglogenesis in tissues of patients at risk of coronary artery disease  
CC or ischemic tissue damage.  
CC note: this sequence does not appear in the specification; it was created  
CC using information provided.  
XX  
SQ Sequence 669 AA:  
  
Query Match 30.3%; Score 495; DB 21; Length 669;  
Best Local Similarity 46.2%; Pred. No. 2,4e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;  
  
OY 11 NPELREKSRDAARSRQSETEVLYQLAHTLPFAGVSAHLDKASTIMRLTISYLMNRHC 70  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 14 sseRTkEksrdaArstrsEseVfYelaNqDlRlphnvshldkaSwmrltIsyIvvrKl1 73  
OY 71 AAGGKRGATGRILDEPGGFRNGTHRRGRHGLPYGKCOQA-----PGQSVYDLC 120  
DB 74 dAg-----dlIdedmkkaqmcfyIkaldgfymvltD 105  
OY 121 SSSLIH-----NPPRG-TNFSLELIGHSTFDFTINPCDOEELDALTPRNLSSKKLEAFT 174  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 106 dgdmiYisdvNkYmgltqf--elcgshvfdftphodheememlthnglvkkgqeqt 163  
OY 175 EHNFSLRMKSSTLTSGRNTLNKAATWVKVYHSCGHNMRAYKRPRAOTSPAGSPRSEPRLOCTV 234  
DB ::::::::::::::::::::: :::::::::::::::::::::  
DB 164 gsfIfImkcItIsrgTmnlksatwkvhNctghlhvYdtSnqpCg--yKkprmtctIv 221  
OY 235 LTCEAIP 241  
DB :::::::::::::::::::::  
DB 222 lIcepIp 228  
  
RESULT 10  
AAV84166  
ID AAV84166 standard: Protein; 697 AA.  
XX  
XX AAV84166;  
XX  
XX



XX		03-JUL-2000	(first entry)
DE		A variant of human hypoxia inducible factor-1 alpha protein.	
KW	Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;		
KM	hypoxia inducible gene; hypoxia inducible factor; hypoxia;		
KV	ischemia related damage; angiogenesis; coronary artery disease;		
XX	ischemic tissue damage.		
OS	Synthetic.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	Misc-difference 422	/note= "this residue is optionally not Ser, and is preferably Gly"	
FT	Misc-difference 423	/note= "this residue is optionally not Thr, and is preferably Ala"	
XX	WO200010578-A1.		
PN	02-MAR-2000.		
PD			
PF	25-AUG-1999; 99WO-US19416.		
XX	25-AUG-1998; 98US-0148547.		
PR	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
XX			
PA	Semenza GL;		
PI			
DR	WPI; 2000-246493/21.		
XX			
PT	Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for treating hypoxia or ischemia-related tissue damage -		
PS	Claim 1; Page -: 96pp; English.		
CC	The present sequence represents a variant of hypoxia-inducible factor (HIF)-1 alpha, comprising amino acids 1-391 and 521-826 of the wild type protein (see AY6407). The HIF-1alpha variants are stable under hypoxic and non-hypoxic conditions. The variants comprises amino acid residues 1-391 and 521-826, 549-826, 576-826, 469-826, 494-826, 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha polypeptide, in which residues 551 and 552 are not serine and threonine, respectively. The HIF-1alpha variant polynucleotide sequences are useful for increasing expression of a hypoxia inducible gene in a cell. They is also useful for providing constitutive expression of a hypoxia inducible factor in a cell, and for reducing or preventing hypoxia or ischemia related damage. The variant HIF-1alpha polypeptides are useful for providing prophylactic therapy for inducing the level of angiogenesis in tissues of patients at risk of coronary artery disease or ischemic tissue damage.		
CC	note: this sequence does not appear in the specification; it was created using information provided.		
CC			
XX	Sequence 697 AA;		
SO			
QY	Query Match	30.3%; Score 495; DB 21; Length 697;	
	Best Local Similarity	46.2%; Pred. No. 2.5e-42;	
	Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;		
DQ	11 NTELRKEKSROAARRRRSODETVLQLNHTLPFARGVSANHLDKASIMRLTISYLMNRHC 70 :::     :     :     :     :     :     :     :     :     :     :     :		
DQ	14 sserfkeksrdaatsrtkesevfyelahnqilprhnvshldkasvmrltitsylylvrkll 73    :     :     :     :     :     :     :     :     :     :     :		
OY	71 AAGKGKRGATGRLLEPGSGGFRTGNRGRNGCLPYGKCQQA-----PGSYDLC 120    :     :     :     :     :     :     :     :     :     :     :		
DB	74 dag-----dltdedmkqgmcfykaldgfmwytld 105    :     :     :     :     :     :     :     :     :     :     :		
OY	121 SSLSLIH-----NPPRG-TNFSELIGHSIFDFIHPCDDEELODALTPRNLSKKILEAPT 174		

Db	106	dgdmiysadnvnkymgtlqf--eltghsvdfdcphcdheemremllthnrgllyvkkqegnt	163
Qy	175	ERHPSLRKSTSTSRGRTLNLKATWNVLCSCGHRAYKPPAQTSPAGSRSEPPQLCLV	234
Db	164	grsfllmckcltsrgtltmiskstkwvlhctghlyhvdtnsnqpqcg--ykkppmtclv	221
Qy	235	LICERAP 241	
Db	222	liceplp 228	
RESULT 11			
AAV84173	AAV84173 standard; Protein; 701 AA.		
XX	AAV84173:		
XX	03-JUL-2000	(first entry)	
XX	A variant of human hypoxia inducible factor-1 alpha protein.		
DE	Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;		
KW	hypoxia inducible gene; hypoxia inducible factor; hypoxia;		
KW	ischemia related damage; angiogenesis; coronary artery disease;		
KW	ischemic tissue damage.		
XX	Synthetic.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FH	Misc-difference 426	/note= "this residue is not Ser, and is	
FT	FT	preferably Gly"	
FT	Misc-difference 427	/note= "this residue is not Thr, and is	
FT	FT	preferably Ala	
FN	WC200010578-A1.		
XX	02-MAR-2000.		
XX	25-AUG-1999;	99WO-US19416.	
PF	25-AUG-1998;	98US-0148547.	
XX	(UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
PA	Semenza GL;		
P1	WPI; 2000-246493/21.		
DR	Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for		
PT	treating hypoxia or ischemia-related tissue damage -		
XX	Claim 1; Page -: 96pp; English.		
XX	The present sequence represents a variant of hypoxia-inducible factor		
CC	(HIF)-1 alpha, comprising amino acids 1-391 and 517-826 of the wild		
CC	type protein (see AAY69407). The HIF-1alpha variants are stable under		
CC	hypoxic and non-hypoxic conditions. The variants comprises amino acid		
CC	residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826,		
CC	508-826, 512-826 or 517-826 of the wild type human HIF-1alpha		
CC	polypeptide, in which residues 551 and 552 are not serine and threonine,		
CC	respectively. The HIF-1alpha variant polynucleotide sequences are useful		
CC	for increasing expression of a hypoxia inducible gene in a cell. They		
CC	is also useful for providing constitutive expression of a hypoxia		
CC	inducible factor in a cell, and for reducing or preventing hypoxia or		
CC	ischemia related damage. The variant HIF-1alpha polypeptides are useful		
CC	for providing prophylactic therapy for inducing the level of		
CC	angiogenesis in tissues of patients at risk of coronary artery disease		
CC	or ischemic tissue damage.		
CC	note: this sequence does not appear in the specification: it was created		

CC using information provided.  
XX

Sequence 701 AA:

Query Match 30.3%; Score 495; DB 21; Length 701;  
Best Local Similarity 46.2%; Pred. No. 2.5e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

QY 11 NTELRKESDAARRSRSEVLYOLAHLPARGVSAHLDRKASIMRLTSTYLRMRIC 70  
DB 14 sserikksrdaarsrskesevfyelabqplphnvshldkasmrltistylrvkll 73  
QY 71 AAGGKRGRATGRLLPEGGGFRHGTNRGRHGLPVGKCOQA-----PGQSYDLC 120  
DB 74 dag-----dlidddmkagmcfylkaldfvmvltld 105  
QY 121 SSSLIIH-----NPPRG-TNSLELIGHSTFDTHPCQDELDALTPRPLSKKKLEAPT 174  
DB 106 dgdmllyisdhvnkymglgtf--eltghsvldftpcdheemremllthnglvkkyeqnt 163  
QY 175 ERHPSLRKSTLTSGRTLNKATWVLHCSGHRAYKPPAQTSPAGSPRSEPPLOCIV 234  
DB 164 grefllmktcltsrgtmtlksatwkvlnctghihvtdnsgpqcg--ykkpmtciv 221  
QY 235 LICEAIP 241  
DB 222 lliceip 228

RESULT 12

AAy84172 standard; Protein; 710 AA.

AC AAY84172;

DT 03-JUL-2000 (first entry)

DE A variant of human hypoxia inducible factor-1 alpha protein.

XX Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
KW ischemia related damage; angiogenesis; coronary artery disease;  
KW ischemic tissue damage.

OS Synthetic.  
OS Homo sapiens.

PH Key Location/Qualifiers

FT Misc-difference 435 /note- "this residue is not Ser, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

FT Misc-difference 436 /note- "this residue is not Thr, and is

PS Claim 1; Page -: 96pp; English.

XX The present sequence represents a variant of hypoxia-inducible factor  
CC (HIF)-1 alpha, comprising amino acids 1-391 and 508-826 of the wild  
CC type protein (see AAY84171). The HIF-1alpha variants are stable under  
CC hypoxic and non-hypoxic conditions. The variants comprises amino acid  
CC residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826,  
CC 508-826, 512-826 or 517-826 of the wild type human HIF-1alpha  
CC polypeptide, in which residues 551 and 552 are not serine and threonine,  
CC respectively. The HIF-1alpha variant polynucleotide sequences are useful  
CC for increasing expression of a hypoxia inducible gene in a cell. They  
CC are also useful for providing constitutive expression of a hypoxia  
CC inducible factor in a cell, and for reducing or preventing hypoxia or  
CC ischemia related damage. The variant HIF-1alpha polypeptides are useful  
CC for providing prophylactic therapy for inducing the level of  
CC angiogenesis in tissues of patients at risk of coronary artery disease  
CC or ischemic tissue damage.  
CC note: this sequence does not appear in the specification; it was created  
CC using information provided.

Sequence 710 AA:

Query Match 30.3%; Score 495; DB 21; Length 710;  
Best Local Similarity 46.2%; Pred. No. 2.6e-42;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

QY 11 NTELRKESDAARRSRSEVLYOLAHLPARGVSAHLDRKASIMRLTSTYLRMRIC 70  
DB 14 sserikksrdaarsrskesevfyelabqplphnvshldkasmrltistylrvkll 73  
QY 71 AAGGKRGRATGRLLPEGGGFRHGTNRGRHGLPVGKCOQA-----PGQSYDLC 120  
DB 74 dag-----dlidddmkagmcfylkaldfvmvltld 105  
QY 121 SSSLIIH-----NPPRG-TNSLELIGHSTFDTHPCQDELDALTPRPLSKKKLEAPT 174  
DB 106 dgdmllyisdhvnkymglgtf--eltghsvldftpcdheemremllthnglvkkyeqnt 163  
QY 175 ERHPSLRKSTLTSGRTLNKATWVLHCSGHRAYKPPAQTSPAGSPRSEPPLOCIV 234  
DB 164 grefllmktcltsrgtmtlksatwkvlnctghihvtdnsgpqcg--ykkpmtciv 221  
QY 235 LICEAIP 241  
DB 222 lliceip 228

RESULT 13

AAy84171 standard; Protein; 724 AA.

AC AAY84171;

DT 03-JUL-2000 (first entry)

DE A variant of human hypoxia inducible factor-1 alpha protein.

XX Human: hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
KW ischemia related damage; angiogenesis; coronary artery disease;  
KW ischemic tissue damage.

OS Synthetic.  
OS Homo sapiens.

PH Key Location/Qualifiers

FT Misc-difference 449 /note- "this residue is not Ser, and is

FT Misc-difference 450 /note- "this residue is not Thr, and is

FT Misc-difference 450 /note- "this residue is not Thr, and is



Db	106	dgdmyjisdvnykymglqf--eltgysvdfdtbpcdhneemrlthnrglvkkykqent	163
Oy	175	ERHPSLRKRSRLTSGRGLRLNKAATKWKVLRHCSGMRAKRPQAOTSPAGSRSEPPLOCIV	234
Db	164	qrsflrmkcrltsgrtgmiksatwkvlnctghlhyvdtltnsqpcq--ykkpmtclv	221
Oy	235	LICEAIR 241	
Db	222	liceip 228	
RESULT 15			
ID	AA84169		
ID	AA84169	standard; Protein; 789 AA.	
XX	AA84169:		
AC			
XX			
DT	03-JUL-2000	(first entry)	
XX			
DE	A variant of human hypoxia inducible factor-1 alpha protein.		
XX			
KW	Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;		
KW	hypoxia inducible gene; hypoxia inducible factor; hypoxia;		
KW	ischemia related damage; angiogenesis; coronary artery disease;		
KW	ischemic tissue damage.		
XX			
OS	Synthetic.		
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	Misc-difference 514	/note= "this residue is not Ser, and is	
FT		preferably Gly"	
FT	Misc-difference 515	/note= "this residue is not Thr, and is	
FT		preferably Ala"	
FT	Misc-difference 515		
XX			
PN	WO200010578-A1.		
XX			
PD	02-MAR-2000.		
XX			
PF	25-AUG-1999;	99WO-US19416.	
XX			
PR	25-AUG-1998;	98US-0148547.	
XX			
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
XX			
PI	Semenza GU;		
XX			
DR	WPI; 2000-246493/21.		
XX			
PT	Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for		
PT	treating hypoxia or ischemia-related tissue damage -		
XX			
PS	Claim 1; Page -: 96pp; English.		
XX			
CC	The present sequence represents a variant of hypoxia-inducible factor		
CC	(HIF)-1 alpha, comprising amino acids 1-391 and 429-826 of the wild		
CC	type protein (see AA68407). The HIF-1alpha variants are stable under		
CC	hypoxic and non-hypoxic conditions. The variants comprises amino acid		
CC	residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826,		
CC	508-826, 512-826 or 517-826 of the wild type human HIF-1alpha		
CC	polypeptide, in which residues 551 and 552 are not serine and threonine,		
CC	respectively. The HIF-1alpha variant polynucleotide sequences are useful		
CC	for increasing expression of a hypoxia inducible gene in a cell. They		
CC	is also useful for providing constitutive expression of a hypoxia		
CC	inducible factor in a cell, and for reducing or preventing hypoxia or		
CC	ischemia related damage. The variant HIF-1alpha polypeptides are useful		
CC	for providing prophylactic therapy for inducing the level of		
CC	angiogenesis in tissues of patients at risk of coronary artery disease		
CC	or ischemic tissue damage.		
CC	note: this sequence does not appear in the specification; it was created		

```
CC      using information provided.
xx
SQ      Sequence       789 AA;

Query Match          30.3%; Score 495; DB 21; Length 789;
Best Local Similarity 46.2%; Pred No. 3e-42;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6

QY      11 NTELREKSRDARSRSOETEVLYOLATTLPPFANGVSAHLRKASIMRLTISYLPMHRLC 70
        ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      14 sserirkeksridaarrriserevefyelahnqlrlphnvshldkaavmrlltisylrvkl 73
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      71 AAGGKRGRATGRLPLEGGGFRRHGTHRRRGHLPGVKCOA-----DGPOSVDLC 120
        || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      74 dag-----dlidiedmkaqmctfylkaldfymvltld 105
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      121 SSSLH-----NPTRG-TNFSLLEIGHSTFDRIHQDOBELODALTTPRNLSKKLLEAPT 174
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      106 dgdmyisdnhvkmgjltqf--elcgshvfdfthpdcbeamremlthrnglvkksgengt 163
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      175 EHNFLRMKSTLTTSNGRTINLKAAITWKVLHGSGMHRAYRPRPQTSPAGSRRSEPLQCIV 234
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      164 gsfiflmkciltstpyrtmlklsatwvlhctgfhlyvdtnsqpqcg--ytkkpmtictv 221
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      235 LICEAIP 241
        ||||| |
DB      222 liceaip 228
```

Search completed: September 8, 2002, 07:33:25  
Job time: 5683 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 8, 2002, 06:45:32 : Search time 19.62 Seconds  
(without alignments)  
382.195 Million cell updates/sec

Title: US-09-896-791b-3  
Perfect score: 1636  
Sequence: 1 MALGIGRVNSNTELRKEKSR.....TESSLPSTWVLMALRNKCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCYUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	495	30.3	826	1 US-08-785-241-6	Sequence 6, Appl1
2	495	30.3	826	2 US-08-480-473B-2	Sequence 2, Appl1
3	495	30.3	826	3 US-08-915-213-2	Sequence 2, Appl1
4	495	30.3	826	3 US-09-148-547-2	Sequence 2, Appl1
5	495	30.3	826	4 US-09-235-217-2	Sequence 2, Appl1
6	495	30.3	826	5 PCT-US96-10251-2	Sequence 2, Appl1
7	489	29.9	810	1 US-08-785-241-7	Sequence 7, Appl1
8	477.5	29.2	875	1 US-08-785-241-5	Sequence 5, Appl1
9	466	28.5	870	1 US-08-785-241-4	Sequence 4, Appl1
10	444	27.1	373	2 US-08-480-473B-3	Sequence 3, Appl1
11	444	27.1	373	3 US-08-915-213-3	Sequence 3, Appl1
12	444	27.1	373	4 US-09-235-217-3	Sequence 3, Appl1
13	444	27.1	373	5 PCT-US96-10251-3	Sequence 3, Appl1
14	444	27.1	805	2 US-08-480-473B-4	Sequence 4, Appl1
15	444	27.1	805	3 US-08-915-213-4	Sequence 4, Appl1
16	444	27.1	805	4 US-09-235-217-4	Sequence 4, Appl1
17	444	27.1	805	5 PCT-US96-10251-4	Sequence 4, Appl1
18	294	18.0	594	2 US-08-785-310A-5	Sequence 5, Appl1
19	293.5	17.9	594	2 US-08-785-310A-6	Sequence 6, Appl1
20	203	12.4	54	2 US-08-480-473B-36	Sequence 36, Appl1
21	203	12.4	54	2 US-08-915-213-36	Sequence 36, Appl1
22	203	12.4	54	4 US-09-235-217-36	Sequence 36, Appl1
23	197	12.0	50	4 US-09-150-460B-17	Sequence 17, Appl1
24	188	11.5	50	4 US-09-150-460B-18	Sequence 18, Appl1
25	142.5	8.7	51	4 US-09-150-460B-15	Sequence 15, Appl1
26	136.5	8.3	789	4 US-08-971-188-9	Sequence 9, Appl1
27	131	8.0	53	2 US-08-480-473B-38	Sequence 38, Appl1

28	131	8.0	53	3 US-08-915-213-38	Sequence 38, Appl1
29	131	8.0	53	4 US-09-235-217-38	Sequence 38, Appl1
30	125.5	7.7	626	4 US-08-971-188-10	Sequence 10, Appl1
31	124.5	7.6	51	4 US-09-150-460B-16	Sequence 16, Appl1
32	120.5	7.4	631	4 US-08-971-188-8	Sequence 8, Appl1
33	107	6.5	102	2 US-08-480-473B-47	Sequence 47, Appl1
34	107	6.5	102	3 US-08-915-213-47	Sequence 47, Appl1
35	107	6.5	102	4 US-09-235-217-47	Sequence 47, Appl1
36	106	6.5	808	4 US-08-971-188-12	Sequence 12, Appl1
37	106	6.5	848	1 US-08-045-806-4	Sequence 4, Appl1
38	106	6.5	848	1 US-08-366-051B-4	Sequence 4, Appl1
39	103	6.3	716	4 US-08-971-188-5	Sequence 5, Appl1
40	102	6.2	716	4 US-08-971-188-11	Sequence 11, Appl1
41	102	6.2	737	4 US-08-971-188-4	Sequence 4, Appl1
42	99	6.1	501	2 US-08-660-963-13	Sequence 13, Appl1
43	91.5	5.6	846	3 US-08-885-291-55	Sequence 55, Appl1
44	91.5	5.6	846	3 US-09-107-847-2	Sequence 2, Appl1
45	91.5	5.6	846	4 US-09-496-672-55	Sequence 55, Appl1

ALIGNMENTS

```
RESULT 1
US-08-785-241-6
: Sequence 6, Application US/08785241
: Patent No. 5695963
:
: GENERAL INFORMATION:
:   APPLICANT: McKnight, Steven L.
:   APPLICANT: Russell, David W.
:   APPLICANT: Tian, Hui
:   TITLE OF INVENTION: Endothelial PAS Domain Protein
:   NUMBER OF SEQUENCES: 7
:   CORRESPONDENCE ADDRESS:
:   ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
:   STREET: 268 BUSH STREET, SUITE 3200
:   CITY: SAN FRANCISCO
:   STATE: CALIFORNIA
:   COUNTRY: USA
:   ZIP: 94104
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/785,241
:   FILING DATE: 17-JAN-1997
:   CLASSIFICATION: 435
:   ATTORNEY/AGENT INFORMATION:
:   NAME: OSMAN, RICHARD A
:   REGISTRATION NUMBER: 36,627
:   REFERENCE/DOCKET NUMBER: UTSd:1229
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (415) 343-4341
:   TELEFAX: (415) 343-4342
:   INFORMATION FOR SEO ID NO: 6:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 826 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: peptide
:   US-08-785-241-6
```

Query Match 30.3% Score 495; DB 1; Length 826;  
Best Local Similarity 46.2% Pred. No. 1.1e-45;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;  
OY 11 NTELRKEKSRPAASRRQETFEVLYOLAHNTLPFARGVSAHLDDKASIMPLTISYLRMRKLC 70  
DB 14 SSRRKRKSRPAASRRKSESEVFETLAHQPLPLPHNVSYHLDDKASVMTLTISYLRVRKLL 73

```
OY 71 AAGKRGATGRLLPEPGGFRHGRHGLPVGKCOA-----PGQSYDLC 120
DB 74 DAG-----DLDIEDMKRQNMCFYLKALDGFVAVLTD 105
OY 121 SSSLIH-----NPTPG-TNFSLELIGHSIDFTHPCDOEELQDALTPRPNSKKLEAPT 174
DB 106 DGDMIYISDNVKNKYMGLTF--ELTGHSVFEDTHPCDHEEREMLTTHRNGLVKKGEQNT 163
OY 175 ERHSLRMRKSTLTSRGRTLNKATWVLCSCGHMRAKPPAOTSPPAGSPRSEPPLOCLV 234
DB 164 QRSFELMRKCTLTSRGRTLNKATWVLCSCGHMRAKPPAOTSPPAGSPRSEPPLOCLV 221
OY 235 LICEAIP 241
DB 222 LICEPIR 228

RESULT 2
US-08-480-473b-2
; Sequence 2, Application US/08480473b
; Patent No. 5882914
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,473b
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-473b-2

Query Match 30.3%; Score 495; DB 2; Length 826;
Best Local Similarity 46.2%; Pred. No. 1.le-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
OY 175 ERHSLRMRKSTLTSRGRTLNKATWVLCSCGHMRAKPPAOTSPPAGSPRSEPPLOCLV 234
DB 164 QRSFELMRKCTLTSRGRTLNKATWVLCSCGHMRAKPPAOTSPPAGSPRSEPPLOCLV 221
OY 235 LICEAIP 241
DB 222 LICEPIR 228

RESULT 3
US-08-915-213-2
; Sequence 2, Application US/08915213
; Patent No. 6020462
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,213
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,473
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-915-213-2

Query Match 30.3%; Score 495; DB 3; Length 826;
Best Local Similarity 46.2%; Pred. No. 1.le-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```



QY	235	LICEAIP	241
Db	222	LICEPIP	228

```

RESULT      4
US-09-148-547-2
; Sequence 2, Application US/09148547
; Patent No. 6124131
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use
; FILE REFERENCE: 07265/151001
; CURRENT APPLICATION NUMBER: US/09/148,547
; CURRENT FILING DATE: 1998-06-25
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-148-547-2

```

Query Match	30.3%	Score 495	DB 3	length 826
Best Local Similarity	46.2%	Pred. No.	1,1e-45	
Matches 114	Conservative 26	Mismatches 59	Indels 48	Gaps 6

[illegible]

QY	235	LICEAIP	241
Db	222	LICEPIP	228

RESULT 5  
US-09-235-217-2  
Sequence 2, Application US/09235217  
Patent No. 6222018  
GENERAL INFORMATION:  
APPLICANT: Semenza, Gregg L.  
TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/235, 217  
FILING DATE:  
CLASSIFICATION:

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER:  US 08/480,473
3      FILING DATE:
4      ATTORNEY/AGENT INFORMATION:
5      NAME:  Halle, Lisa A.
6      REGISTRATION NUMBER:  38,347
7      REFERENCE/DOCKET NUMBER:  07265/053001
8      TELECOMMUNICATION INFORMATION:
9      TELEPHONE:  619/678-5070
10     TELEFAX:  619/678-5099
11     INFORMATION FOR SEQ ID NO:  2:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH:  826 amino acids
14     TYPE:  amino acid
15     TOPOLOGY:  linear
16     MOLECULE TYPE:  protein
17     US-09-235-217-2

```

Query Match	30.38;	Score	495;	DB 4;	Length	826;			
Best Local Similarity	46.28;	Pred. No.	1.1e-45;						
Matches	114;	Conservative	26;	Mismatches	59;	Indels	48;	Gaps	6

```

OY 11 NTELEKESRPAASRSROEVEVYLOLAHLPFARGSAHLIDKRSIMRTSYRYMRKL 70
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 SSEKREKESRPAASRSRKESEVEVELAHOLPRLHNWSSHLIDKASVNRILTSTYLRVKLL 73
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 71 AAGKRGKATGRLLPEGPGRHGHTRGRHGLPEVKGCOQA-----PGQSVYDLG 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 DAG-----DLIEDDKMQNMCFYLKALDGFVNYLTD 105
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 121 SSSLIH-----NPPRG-TNFSLELIGHSPFHPRCOEELQDLATLTPRPLSKKLEAPT 174
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 DGDMYTISDNVKNKVMGLQOF--ELTGHVSFDPETHPCDHEEKREMLTRHNGJLVKKGKEDONT 163
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 ERHFSLRMKSTLTSGKRTLTNKAATNMVTLHCSGMRARYKPPAONSPAGSPRSEPPDCLCY 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 QRSFPLLRKCTVLSRGKTRMINKSATKVMVLCHTGHIHYDTNSNOPCC--TKKPPMTCV 222
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY	235	LICEAIP	241
Db	222	LICEPIP	228

RESULT 6  
PCT-US96-10251-2  
Sequence 2, Application PC/TUS9610251  
GENERAL INFORMATION:  
APPLICANT: The Johns Hopkins University School of Medicine  
TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10251  
FILING DATE: 06-JUN-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/053w01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070



Query Match 29.2%; Score 477.5; DB 1; Length 875;  
Best Local Similarity 45.3%; Pred. No. 1e-43; Indels 27; Gaps 9;  
Matches 115; Conservative 33; Mismatches 79; Indels 27; Gaps 9;

QY 1 MALGIRVSNTELREKRSDDARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
DB 1 MADRKKRKSSELRKREKRDAAACRKRKETEYFTELAEHLPLPHYSVSHLDKASIMRLA 60  
QY 61 ISYLRMHL-----CAAGKRGATGRLLEPGGFRHGRHGRHGLPVGKCOQAPGPOS 116  
DB 61 ISFLRTHKLLSSVCSNESAADQOM-----DNLTKALEGF-----IAYVTD 105  
QY 117 VLCSSSLHNPPTGPNFSLDELGHISFDFIHPCDDELDALTPR--PULSKKLEAPY 174  
DB 106 GMIFLSENIKRFMGLT-OVELTGHISIFDFTHPCHDEIRENLSLKGSGFGKSKDVST 164  
QY 175 ERHFSLRMSTLTSRGTNLKATWK-VLHCSGHRAYKP-PAGTSPAGSPRSEPLOC 232  
DB 165 ERDFEHRMCTVYNNRGTYNLKSATYKSVLHCTGOVRYVNNCPHNSLCSG--KEPLSLC 222

QY 233 VLICEAIPOLPFR 246  
DB 223 LIMCEPI-QHPSH 235

RESULT 9  
US-08-785-241-4  
; Sequence 4, Application US/08785241  
; Patent No. 5695963  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; APPLICANT: Tian, Hui  
; TITLE OF INVENTION: Endothelial PAS Domain Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,241  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1229  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4341  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 870 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-785-241-4

Query Match 28.5%; Score 466; DB 1; Length 870;  
Best Local Similarity 43.9%; Pred. No. 1.8e-42;  
Matches 111; Conservative 34; Mismatches 82; Indels 26; Gaps 8;

QY 1 MALGIRVSNTELREKRSDDARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60

DB 1 MADRKKRKSSELRKREKRSDDAARCRKRKETEYFTELAEHLPLPHYSVSHLDKASIMRL 60  
QY 61 ISYLRMHL-----CAAGKRGATGRLLEPGGFRHGRHGRHGLPVGKCOQAPGPOS 116  
DB 61 ISFLRTHKLLSSVCSNESAADQOM-----DNLTKALEGF-----IAYVTD 105  
QY 117 VLCSSSLHNPPTGPNFSLDELGHISFDFIHPCDDELDALTPR--PULSKKLEAPY 174  
DB 106 GMIFLSENIKRFMGLT-OVELTGHISIFDFTHPCHDEIRENLSLKGSGFGKSKDMST 164  
QY 175 ERHFSLRMSTLTSRGTNLKATWKVLCGSHMRAYKP-PAGTSPAGSPRSEPLOC 233  
DB 165 ERDFEHRMCTVYNNRGTYNLKSATYKSVLHCTGOVRYVNNCPHNSLCSG--YKPLSLC 222

QY 234 VLICEAIPOLPFR 246  
DB 223 LIMCEPI-QHPSH 234

RESULT 10  
US-08-480-473B-3  
; Sequence 3, Application US/08480473B  
; Patent No. 5882914  
; GENERAL INFORMATION:  
; APPLICANT: Semenza, Gregg L.  
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,473B  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hallie, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/053001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 373 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-473B-3

Query Match 27.1%; Score 444; DB 2; Length 373;  
Best Local Similarity 44.3%; Pred. No. 1.4e-40;  
Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6;

QY 23 ARSRROETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHLCAAGKRGATGR 82  
DB 5 AGSRKRSSEYFTELAEHLPLPHVSSHLDKASVMLRTISYLRKLLDAG----- 55  
QY 83 LPEGGGFRHGRHGRHGLPVGKCOQA-----PGOSVDCSSSLH-----N 127  
DB 56 -----DLIEDDKAKQMNCFYALKALDGFVWVLLDDDMITISDNVN 96

OY 128 PTPG-TNFSLEIGHSIDFTHPCOELOALTPRPNSKTKLAPTERHSLRMKSTL 186  
 Db 97 KYMGLTGF--ELTGHVSDFTHPCDHEEMRLTNRNGLVKKGKQONTORSFLLRMKCTL 154  
 OY 187 TSGRTNLKATWKLVLHCSGHMRAVKKPAOTSPPAGSPRSEPPLOCLVICAIP 241  
 Db 155 TSGRTNMIKATWKLVLHCTGHIHYDINSNOPOCG--YKPPMCLVLCIP 207

RESULT 11  
 US-08-915-213-3  
 ; Sequence 3, Application US/08915213  
 ; Patent No. 6020462

; GENERAL INFORMATION:  
 ; APPLICANT: Semenza, Gregg L.  
 ; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/915,213  
 ; FILING DATE: 20-AUG-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,473  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 07265/053001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 373 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-915-213-3

Query Match 27.1%; Score 444; DB 3; Length 373;  
 Best Local Similarity 44.3%; Pred. No. 1.4e-40;  
 Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6;

OY 23 AASRSQETEVYQLAHTLPFARGVSAHLDRKASIMRLTISYLRMRLCAAGKRGRRATGR 82  
 Db 5 AGSRKSESEVEYELAHQPLPHNVSSHLDKASVMRLTISYLRVKRLDAG----- 55  
 OY 83 LLPEPGGFRHGTNRGRHGLPVGKCOA-----PGQSYDLCSSLIH-----N 127  
 Db 56 -----DLIEDMKAKQNCFLKALDGFVWVLTJDDGMITYSDNVN 96  
 OY 128 PTPG-TNFSLEIGHSIDFTHPCOELOALTPRPNSKTKLAPTERHSLRMKSTL 186  
 Db 97 KYMGLTGF--ELTGHVSDFTHPCDHEEMRLTNRNGLVKKGKQONTORSFLLRMKCTL 154  
 OY 187 TSGRTNLKATWKLVLHCSGHMRAVKKPAOTSPPAGSPRSEPPLOCLVICAIP 241  
 Db 155 TSGRTNMIKATWKLVLHCTGHIHYDINSNOPOCG--YKPPMCLVLCIP 207

RESULT 12  
 US-09-235-217-3  
 ; Sequence 3, Application US/09235217  
 ; Patent No. 622018

; GENERAL INFORMATION:  
 ; APPLICANT: Semenza, Gregg L.  
 ; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/235,217  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,473  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 07265/053001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 373 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-235-217-3

Query Match 27.1%; Score 444; DB 4; Length 373;  
 Best Local Similarity 44.3%; Pred. No. 1.4e-40;  
 Matches 104; Conservative 24; Mismatches 59; Indels 48; Gaps 6;

OY 23 AASRSQETEVYQLAHTLPFARGVSAHLDRKASIMRLTISYLRMRLCAAGKRGRRATGR 82  
 Db 5 AGSRKSESEVEYELAHQPLPHNVSSHLDKASVMRLTISYLRVKRLDAG----- 55  
 OY 83 LLPEPGGFRHGTNRGRHGLPVGKCOA-----PGQSYDLCSSLIH-----N 127  
 Db 56 -----DLIEDMKAKQNCFLKALDGFVWVLTJDDGMITYSDNVN 96  
 OY 128 PTPG-TNFSLEIGHSIDFTHPCOELOALTPRPNSKTKLAPTERHSLRMKSTL 186  
 Db 97 KYMGLTGF--ELTGHVSDFTHPCDHEEMRLTNRNGLVKKGKQONTORSFLLRMKCTL 154  
 OY 187 TSGRTNLKATWKLVLHCSGHMRAVKKPAOTSPPAGSPRSEPPLOCLVICAIP 241  
 Db 155 TSGRTNMIKATWKLVLHCTGHIHYDINSNOPOCG--YKPPMCLVLCIP 207

RESULT 13  
 PCT-US96-10251-3  
 ; Sequence 3, Application PC/TUS9610251  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Johns Hopkins University School of Medicine  
 ; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 ; NUMBER OF SEQUENCES: 35





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:32:32 : Search time 37.72 Seconds  
(without alignments)  
782.063 Million cell updates/sec

Title: US-09-896-791b-3  
Perfect score: 1636  
Sequence: 1 MALGQVRNSNTLREKRSR.....TESSLPSVLMALNRKNC PG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802	49.0	667	2 JC7771	hypoxia inducible
2	503	30.7	811	2 JC7619	hypoxia-inducible
3	495	30.3	826	2 I38972	hypoxia-inducible
4	492	30.1	810	2 JC4837	hypoxia-inducible
5	485	29.6	813	2 JC5809	hypoxia-inducible
6	281	17.2	1505	2 JC4851	hypoxia-inducible
7	256	15.6	248	2 A58520	single-minded gene
8	202	12.3	655	2 A29945	neurogenesis regul
9	169.5	10.4	823	2 T21943	hypothetical prote
10	169.5	10.4	825	2 T21944	hypothetical prote
11	142	8.7	805	2 JC7635	aryl hydrocarbon r
12	138.5	8.5	392	2 JC7633	aryl hydrocarbon n
13	138.5	8.5	791	2 A56241	aryl hydrocarbon r
14	136.5	8.3	776	2 A55448	Ah receptor nuclea
15	136.5	8.3	789	2 I59550	aryl hydrocarbon r
16	129.5	7.9	358	2 S58376	aryl hydrocarbon r
17	129.5	7.9	451	2 T42397	aryl hydrocarbon r
18	129	7.9	453	2 T19440	hypothetical prote
19	128	7.8	56	2 A41647	neurogenesis regul
20	125.5	7.7	626	2 JC5405	brain and muscle A
21	123.5	7.5	626	2 J60270	Arnt-like PAS prot
22	106	6.5	848	2 S59514	aryl hydrocarbon r
23	104.5	6.4	1059	2 T30557	aryl hydrocarbon r
24	103.5	6.3	258	2 JC5407	brain and muscle A
25	103	6.3	846	2 JC7721	aryl hydrocarbon r
26	103	6.3	1079	2 A70972	probable DNA polym
27	102.5	6.2	716	2 T09462	juvenile hormone r
28	102	6.2	716	2 T09462	juvenile hormone r
29	98	6.0	1058	2 T30556	aryl hydrocarbon r

30	97.5	6.0	326	2 S71755	transcription fact
31	97	5.9	527	2 A75399	hypothetical prote
32	95.5	5.8	920	2 JC7313	aryl hydrocarbon r
33	94.5	5.8	300	2 T24292	hypothetical prote
34	94.5	5.8	1151	2 T24541	hypothetical prote
35	93.5	5.7	634	2 T00359	hypothetical prote
36	92.5	5.7	1751	2 T09394	gag-pro-pol polypr
37	90	5.5	743	2 G83726	assimilatory nitra
38	90	5.5	1920	2 T13893	gene hindsight pro
39	89.5	5.5	1891	2 T13594	hypothetical prote
40	89	5.4	853	2 S38375	aryl hydrocarbon r
41	89	5.4	3133	2 S52093	hemocytin - silkwo
42	87.5	5.3	647	2 S06450	steroid hormone re
43	87.5	5.3	805	2 A46266	aryl hydrocarbon r
44	87.5	5.3	2481	2 A43908	fibronectin - Afri
45	87	5.3	1140	2 B70729	hypothetical prote

ALIGNMENTS

RESULT 1  
JC7771  
hypoxia inducible factor-3 alpha - human  
C:Species: Homo sapiens (man)  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: JC7771  
R:Hara, S.; Hamada, J.; Kobayashi, C.; Kondo, Y.; Imura, N.  
Biochem. Biophys. Res. Commun. 287, 808-813, 2001  
A>Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in  
A:Reference number: JC7771; PMID:11573933  
A:Contents: Kidney  
A:Accession: JC7771  
A:Molecule type: mRNA  
A:Residues: 1-667 <HAR>  
A:Cross-References: DBJ:AB054067  
C:Comment: This protein is a heterodimeric transcription factor that belongs to the b  
lved in the regulation of hypoxia-inducible gene expression in human kidney.  
C:Genetics:  
A:Gene: hlf-3alpha  
A:Map position: 19  
A:Keywords: kidney

Query Match 49.0%; Score 802; DB 2; Length 667;  
Best Local Similarity 68.0%; Pred. No. 7.4e-62;  
Matches 172; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

QY 9 RSWTELRKRSRPAARSRSOETEVLYOLAHTLPFARGVSAHDKASIMRLTISYRMHR 68  
|||  
Db 7 RSTTELRKRSRPAARSRSOETEVLYOLAHTLPFARGVSAHDKASIMRLTISYRMHR 66  
|||  
QY 69 LCAAGKRGKRGRLPESGPGFRHGTHRGHGLPVGKCQ-----QAPGPOSV 117  
|||  
Db 67 LCAAG-----EMNOYGAGGELDLACYLKALDGFVWVLAEDPMAY 106  
|||  
QY 118 DLCSSSLINHPPTGNTFS-LELIGHSIFDPFHPCDDELDALTPPNISKKKLEAPTER 176  
|||  
Db 107 -----LSENVSKHLGSLQLELIGHSIFDPFHPCDDELDALTPPOTLSRRKEVAPTER 160  
|||  
QY 177 HPSLRKSTLTSGRITLNKAATWKVLCGSHRAVPAQOTSPPASPSSEPPLOCLTVIA 236  
|||  
Db 161 CFSLRKSTLTSGRITLNKAATWKVLCGSHRAVPAQOTSPPASPSSEPPLOCLTVIA 220  
|||  
QY 237 CEAIPLPFHGA 249  
|||  
Db 221 CEAIPLPFHGA 229  
|||

RESULT 2  
JC7619  
hypoxia-inducible factor 1 alpha - chicken  
C:Species: Gallus gallus (chicken)





RESULT 5  
JC5809  
hypoxia-inducible factor 1 alpha - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 04-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 26-Aug-1999  
C:Accession: JC5809  
R:Ladoux, A.; Frelin, C.  
B:Biochem. Biophys. Res. Commun. 240, 552-556, 1997  
A:Title: Cardiac expressions of HIF-1 alpha and HLF/EPAS, two basic loop helix/PAS domain  
A:Reference number: JC5809; MUID:98065274  
A:Accession: JC5809  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-813 (LAD>  
C:Comment: This protein associates to the aryl hydrocarbon receptor nuclear translocator  
as erythropoietin, vascular endothelial growth factor, the GLUT1 glucose transporter, e  
F:6-144:Region: Basic helix-loop-helix #status predicted

```

Db      67  RNNKKRKKRRDRAACRCRSMETETFMELSAALPLKTDVYNQDLKASVMYITTAFLKIRE 126
      69  L-----CAAGCKRGKRGATGRLLPEGPGGFRHGRHRRNGHLPYGCKCOADGPGVSDYL 119
Qy      127  MLOFVPSLRDNCNDIKODIETAE-----DDQEVKPKLEVETEMLNGAEKREL 174
Db
Qy      120  CSSSL-----IHNPTGTNFS-----LELGHSIFDFIHPDCDEELQDALTPR 162
Db      175  LKQIMDGLLVLSHEGDIITYSENVEYLGIKTDIDLGQDIWEYSHOCDAELKEALSLK 234
      163  PNLSSKKLEAPTE-----RHSLSLPMKSTLTSRGRLLNKAATWKLHCSGHM 209
Qy      235  RELAKQVNDPEQDQNSGVSTHNHRLDFVRLKLTLSRGSINIKASVYVHITHTGL 289
Db

```

```

Db      60 LCDAMQ--PSRAGPLDGVAKELGSHLLQTLDFVFFVASDGMKIMYISETASVHLGLS-- 115
QY      125 INNPRTGTFNPSLELGHSTFDFIHPDQDELDALTPRPNLSSKKL-EAPTEHSELRMK 183
Db      116 -----QVELTGNISIEYIHPSDHDEMTAVLTAHOPLHHLLQEEIEHSFPLRMK 165
QY      184 STLTSGRTLNKAAATWKVLHCSGHRM 210
Db      166 CVLAK--RNAGLTCSGKYVHCSGYIK 190

RESULT  8
A29945
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)
N:Alternate names: single-minded gene protein
C:Species: Drosophila melanogaster
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C:Accession: A29945
R:Crewe, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A>Title: The Drosophila single-minded gene encodes a nuclear protein with sequence siml
A:Reference number: A29945; MUID:88151023
A:Accession: A29945
A:Molecule type: mRNA
A:Residues: 1-655 <CRE>
A:Cross-references: GB:ML9020; NID:q158464; PID:q158465
C:Genetics:
A:Gene: sim
A:Cross-references: FlyBase:FBgn0004666
C:Keywords: DNA binding; transcription regulation

Query Match      12.3%; Score 202; DB 2; Length 655;
Best Local Similarity 28.5%; Pred. No.1.3e-09;
Matches 73; Conservative 35; Mismatches 78; Indels 70; Gaps 10;

```

```

QY      36 QLAHTLPFARGVSAHLDKASIMRLTISYLRMRLCAAGKRGKRGATGRLLPEGPGGFRHGT 95
Db      4 ELAKLPLPAATISQLDKASVIRLTISYLMKRGVFPDG--LGKAKG-----SSPAMRGAT 57
QY      96 HRR-GRHG-----PVGKCOAPQPSVDCSSSLHNPPTGTFNLSLGHST 143
Db      58 IRELGSHELLQTLDFVFFVASDGMKIMYISETASVHLGLS-----QVELTGNISI 105
QY      144 PFIHPDQDELDALTPRPNLSSKKL-----EAPT-----EER 177
Db      106 FEYIHPDQDELDALTPRPNLSSKKL-----EAPT-----EER 177
QY      178 FSLRMKSTLTSGRTLNKAAATWKVLHCSGHRMAYKPPAOTSPAGSPRESEPLQCLVLIC 237
Db      166 FFLRMKCVLAK--RNAGLTSGFKVHCSGYIKARITPPRGDQGS-----LIQNLGYA 218
QY      238 -----EAIPLPFH 246
Db      219 VGHSLPSSAITEIKH 234

RESULT  9
T21943
hypothetical protein F38A6.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21943; T26899
R:Mortimore, B.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z19491
A:Accession: T21943
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-823 <MI2>
A:Cross-references: EMBL:Z92833; PIDN:CA807380.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone F38A6

```

```

R:Ainscough, R.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20282
A:Accession: T26899
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-823 <MI2>
A:Cross-references: EMBL:AL023842; PIDN:CAA19520.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone Y44A6D
C:Genetics:
A:Gene: CESP:F38A6.3a
A:Map position: 5
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 711/1; 768/3

```

```

Query Match      10.4%; Score 169.5; DB 2; Length 823;
Best Local Similarity 26.3%; Pred. No.1.2e-06;
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

QY      9 RSNTELKREKSRDARSRSOETEVLYOLAHTLPFA-RGVSAHLDKASIMRLTISYLRMH 67
Db      110 KRMEKRRKTSRHAARDRSKESDIFDLMKCVPIYEBGTVYHDLRALRVAATICRLR 169
QY      68 RLCA--AGKRGKATGRLLPEGPGGFRHGTGRHGRHGLPVGKCOAPQPSVDCSSSL 124
Db      170 KTAGVLENNLNDNETNEWTED-----TIAECLDG-FVMIVSDSSYL 212
QY      125 INNPRTGTFNPSL--ELIGHSTFDFIHPDQDELDALTPRPNLSSKKLEAP-----T 174
Db      213 YTESVAMYLGLQYDLDLGRALRDLPLPSDYDE-----DKOSKMLKKPGEDTDTT 264
QY      175 ERHFSLRMKSTLTSGRTLNKAAATWK-----VLH-----CSGHRMAYKPPAOTSPAG 222
Db      265 GINMVLRMKTVISPRRCNLKLSALYKSVPLVHKSIVSGHVSFMQ--GITIPAG 318

RESULT  10
T21944
hypothetical protein F38A6.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21944; T26900
R:Mortimore, B.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z19491
A:Accession: T21944
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-825 <MI2>
A:Cross-references: EMBL:Z92833; PIDN:CA807381.1; GSPDB:GN00023; CESP:F38A6.3b
A:Experimental source: clone F38A6
R:Ainscough, R.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z20282
A:Accession: T26900
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-825 <MI2>
A:Cross-references: EMBL:AL023842; PIDN:CAA19521.1; GSPDB:GN00023; CESP:F38A6.3b
A:Experimental source: clone Y44A6D
A:Gene: CESP:F38A6.3b
A:Map position: 5
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 713/1; 770/3

```

```

Query Match      10.4%; Score 169.5; DB 2; Length 825;
Best Local Similarity 26.3%; Pred. No.1.2e-06;
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

QY      9 RSNTELKREKSRDARSRSOETEVLYOLAHTLPFA-RGVSAHLDKASIMRLTISYLRMH 67
Db      110 KRMERRRRTSRHAARDRSKESDIFDLMKCVPIYEBGTVYHDLRALRVAATICRLR 169

```

OY 68 RLCA---AGGKRGATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPQSDVLCSSSL 124  
A:Residues: 1-392 <HSD>  
A:Cross-References: GB:AY007992  
C:Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimeriz  
C:Gene: arntl2x  
DB 170 KTAGVLENNLDNETINTEWTD-----TIAECDDG-FVMITVSDSSL 212  
OY 125 IHNPTGTNFSL---ELIGHSTFDFIHPDOELDALTPRPNLSSKKLEAP-----T 174  
DB 213 YTESVAVYLGITDGLRALRDLHPDSEDEF-----DKQSKMLHKRGEDTDT 264  
OY 175 ERHSLRMRKSTLTSGRFLNLKAAATW---VLH---CSGMRAYKPPAOTSPAG 222  
DB 265 GINWLRMRKTVISPRGCLNLKALYKSVSLYHKSVTSGTHVSEMO-GITIPAG 318  
RESULT 11  
JC7635  
aryl hydrocarbon receptor nuclear translocator 1 - chicken  
N:Alternate names: hypoxia-inducible factor 1 beta  
C:Species: Gallus gallus (chicken)  
C>Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
C:Accession: J07635  
R:Caution, T.; Mendioja, M.A.; Smith, S.M.; Born, J.; Walker, M.K.  
Biochem. Biophys. Res. Commun. 282, 602-607, 2001  
A:Title: Hypoxia regulates avian cardiac Arnt and HIF-1alpha mRNA expression.  
A:Reference number: J07635; MUID:21294777; PMID:11401503  
A:Contents: Embryo  
A:Accession: J07635  
A:Molecule type: mRNA  
A:Residues: 1-805 <CAT>  
C:Comment: This protein, known as hypoxia-inducible factor 1 beta, a member of the basic  
C:Genetics: responds to hormonal and environmental stimuli. It is also a common dimer partner for tran  
A:Gene: arntl  
C:Keywords: transcription factor

Query Match 8.7%; Score 142; DB 2; Length 805;  
Best Local Similarity 20.5%; Pred. No. 0.00028;  
Matches 61; Conservative 43; Mismatches 95; Indels 98; Gaps 8;  
OY 14 LRKESRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMRLCAAG 73  
DB 89 LARENHSEIERRRNRKMTAYITELSDMVPYCSALARKPKDKITILMAVSHM----- 139  
OY 74 GKRGATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPQSDVLCSSSL-----HNPT 129  
DB 140 -KSLGTGTNTSDGTYKPSFLTDQELKHLI---LEADGFLFVYSCETGRVYVSDSVT 194  
OY 130 PGTNF-STELIGHSTFDFIHPDOELDALTPRPN-----LSKKLEAPTERHF 178  
DB 195 PVLNPOQSEWGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKPAAPES-- 252  
OY 179 SLRMRKSTLTSGRFLNLKAAATW----- 201  
DB 253 -----ASKGRITDLKTGTGKKEGQSMRMCGSRSEFICRMRCGNSVDVAVSNRLS 304  
OY 202 -----VLHSGMRAYKPPAOTSPAGSP-RSEPPLOCVLVI 236  
DB 305 FMRNRCRNGLATKDEPHYVYVHCTGYIKAMPAGVSLPDDDPAGQGSKECLVAI 361  
RESULT 12  
JC7633  
aryl hydrocarbon nuclear translocator ARNT2-like factor, ARNT2X - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C>Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
C:Accession: J07633  
R:Hsu, H.J.; Wang, W.D.; Hu, C.H.  
Biochem. Biophys. Res. Commun. 282, 487-492, 2001  
A:Title: Ectopic expression of negative ARNT2 factor disrupts fish development.  
A:Reference number: J07633; MUID:21294759; PMID:11401485  
A:Contents: Heart  
A:Accession: J07633

A:Molecule type: mRNA  
A:Residues: 1-392 <HSD>  
A:Cross-References: GB:AY007992  
C:Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimeriz  
C:Gene: arntl2x  
Query Match 8.5%; Score 138.5; DB 2; Length 392;  
Best Local Similarity 20.8%; Pred. No. 0.00025;  
Matches 55; Conservative 48; Mismatches 103; Indels 59; Gaps 8;  
OY 16 KEKSRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMRLCAAGK 75  
DB 47 RENHSEIERRRNRKMTAYITELSDMVPYCSALARKPKDKITILMAVSHM-----K 96  
OY 76 RGRATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPQSDVLCSSSL-----HNPTG 131  
DB 97 SMRGNTSTDGAYKPSFLTDQELKHLI---LEADGFLFVYVAAETGRVYVSDSVTPV 152  
OY 132 TNF-STELIGHSTFDFIHPDOELDALTPRPNLSSKKL----- 170  
DB 153 LNHPOSEWGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKPAAPES-- 212  
OY 171 -----EAPTRHPSLRMRKSTLTSGRFT---LNLKATWVLHSGMRAY 212  
DB 213 MGSRSFTCRMCGSAPLD-HISLNRSSMRKRYRNGLPKSGEAGVYVHCTGYIKAM 271  
OY 213 KPPAOTSP-AGSPREPPLOCVLVI 236  
DB 272 PPAQMTIPDEDTAGQTSKYCLVAI 296  
RESULT 13  
A56241  
aryl hydrocarbon receptor nuclear translocator protein - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 05-Nov-1999  
C:Accession: A56241  
R:Reisz-Porras, S.; Probst, M.R.; Fukunaga, B.N.; Hankinson, O.  
Mol. Cell. Biol. 14, 6075-6086, 1994  
A:Title: Identification of functional domains of the aryl hydrocarbon receptor nuclea  
A:Reference number: A56241; MUID:94344118  
A:Accession: A56241  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-791 <REI>  
A:Cross-References: GB:U10325; NID:9555687; PIDN:AAA56717.1; PID:9555688  
A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380  
A:Note: authors failed to translate CAG for residue 507 as Gln  
C:Genetics:  
A:Gene: ARNT

Query Match 8.5%; Score 138.5; DB 2; Length 791;  
Best Local Similarity 21.8%; Pred. No. 0.00056;  
Matches 58; Conservative 50; Mismatches 101; Indels 57; Gaps 9;  
OY 14 LRKESRDAARSRSOETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMRLCAAG 73  
DB 89 LARENHSEIERRRNRKMTAYITELSDMVPYCSALARKPKDKITILMAVSHM----- 139  
OY 74 GKRGATGRLLPEGPGFRHGRHGRHGLPVGKCOQAPQSDVLCSSSL-----HNPT 129  
DB 140 -KSLRGNTSTDGAYKPSFLTDQELKHLI---LEADGFLFVYSCETGRVYVSDSVT 194  
OY 130 PGTNF-STELIGHSTFDFIHPDOELDALTPRPN-LSKKLEAP----- 174  
DB 195 PVLNPOQSEWGSTLYDQVHPRDVKLRQLSTSENALTEGKPKCLSKKPAAPES-- 254  
OY 175 -----ERHPSLRMRKSTLTSGRFT---LNLKATWVLHSGMRAY 211  
DB 255 MCGSRSFICRMCGSAPLD-HISLNRSSMRKRYRNGLPKSGEAGVYVHCTGYIKAM 314





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:33:27 ; Search time 19.87 Seconds

(without alignments)  
598.234 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALGLOVRSRNTSLRREKSR.....TESSLPSVLMALNRKNCPC 307

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	495	30.3	822	1	H1FA_MOUSE
2	495	30.3	826	1	H1FA_MOUSE
3	488	29.8	874	1	PAS1_MOUSE
4	467	28.5	870	1	PAS1_MOUSE
5	294	18.0	590	1	NPAL_MOUSE
6	293.5	17.9	594	1	NPAL_MOUSE
7	281	17.2	1507	1	SIWA_MOUSE
8	268	16.4	766	1	SIWA_MOUSE
9	266	16.3	765	1	SIWA_MOUSE
10	265	16.2	657	1	SIWA_MOUSE
11	256	15.6	667	1	SIWA_MOUSE
12	251	15.3	958	1	SIWA_MOUSE
13	249	15.2	673	1	SIWA_MOUSE
14	149	9.1	413	1	CYCL_MOUSE
15	138.5	8.5	791	1	ARNT_MOUSE
16	137.5	8.4	800	1	ARNT_MOUSE
17	136.5	8.3	789	1	ARNT_MOUSE
18	135.5	8.3	790	1	ARNT_MOUSE
19	131.5	8.0	706	1	ARNT_MOUSE
20	131.5	8.0	712	1	ARNT_MOUSE
21	126	7.7	644	1	ARNT_MOUSE
22	122.5	7.5	583	1	BMAL_MOUSE
23	106	6.5	848	1	BMAL_MOUSE
24	91.5	5.6	846	1	BMAL_MOUSE
25	91	5.6	3133	1	BMAL_MOUSE
26	90.5	5.5	805	1	BMAL_MOUSE
27	89.5	5.5	855	1	BMAL_MOUSE
28	89.5	5.5	304	1	HEX1_MOUSE
29	89	5.4	853	1	HEX1_MOUSE
30	87.5	5.3	471	1	HEX1_MOUSE
31	87.5	5.3	587	1	HEX1_MOUSE
32	87.5	5.3	602	1	HEX1_MOUSE
33	87.5	5.3	647	1	HEX1_MOUSE

34	87.5	5.3	2481	1	F1NC_XENLA
35	86.5	5.3	545	1	RTN2_MOUSE
36	86	5.3	700	1	RTN2_MOUSE
37	85	5.2	475	1	3A2_MOUSE
38	84	5.1	213	1	NOUL_MOUSE
39	83	5.1	304	1	HEX1_MOUSE
40	83	5.1	426	1	SMY_MOUSE
41	83	5.1	1024	1	SMY_MOUSE
42	82.5	5.0	649	1	V7OK_MOUSE
43	82.5	5.0	711	1	HEX1_MOUSE
44	82.5	5.0	1159	1	HEX1_MOUSE
45	82.5	5.0	1487	1	ICP4_MOUSE

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	H1FA_MOUSE	495	30.3	822	1	H1FA_MOUSE
AC	061221	061665	061664	008993	008741	
DT	01-NOV-1997	(Rel. 35)	Created			
DT	15-DEC-1998	(Rel. 37)	Last sequence update			
DT	16-OCT-2001	(Rel. 40)	Last annotation update			
DE	Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein).					
DE	H1FA_MOUSE	495	30.3	822	1	H1FA_MOUSE
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=C57BL/6; TISSUE=Hepatocytes;					
RC	MEDLINE=96353491; PubMed=8702901;					
RA	Li H., Ko H.P., Whitlock J.P., Jr.;					
RT	"Induction of phosphoglycerate kinase 1 gene expression by hypoxia. Roles of Arnt and Hif1alpha.";					
RL	J. Biol. Chem. 271:21262-21267(1996).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=129/SV;					
RC	MEDLINE=98034461; PubMed=9368100;					
RA	Luo G., Gu Y.-Z., Jain S., Chan W.K., Carr K.M., Hogenesch J.B., Bradford C.A.;					
RT	"Molecular characterization of the murine Hif-1 alpha locus.";					
RL	Gene Expr. 6:287-299(1997).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=129/SV;					
RC	MEDLINE=97354184; PubMed=9210478;					
RA	Wenger R.H., Rolfs A., Kvietikova I., Spielmann P., Zimmermann D.R., Gassmann M.;					
RT	"The mouse gene for hypoxia-inducible factor-1alpha. Genomic organization, expression and characterization of an alternative first exon and 5' flanking sequence.";					
RL	Eur. J. Biochem. 246:155-165(1997).					
RN	[4]					
RP	SEQUENCE OF 13-822 FROM N.A.					
RC	TISSUE=Hepatocytes;					
RC	MEDLINE=96254028; PubMed=8660378;					
RA	Wenger R.H., Rolfs A., Marti H.H., Guenet J.-L., Gassmann M.;					
RT	"Nucleotide sequence, chromosomal assignment and mRNA expression of mouse hypoxia-inducible factor-1 alpha.";					
RL	Biochem. Biophys. Res. Commun. 223:54-59(1996).					
RN	[5]					
RP	SEQUENCE OF 22-85 FROM N.A.					
RC	TISSUE=Hepatocytes;					
RC	O'Rourke J.F.;					
RT	Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.					
CC	-1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).					
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER					

CC BLH1 PROTEIN, HETERODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.  
CC THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN  
CC DIMERIZE WITH OTHER BLH1-PAS PROTEINS. INTERACTS WITH HSP90 (BY  
CC SIMILARITY)..  
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
CC -1- TISSUE SPECIFICITY: Ubiquitous.  
CC -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY  
CC RESIDE WITHIN THE C-TERMINAL PART.  
CC -1- PMF: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BLH1) FAMILY OF  
CC TRANSCRIPTION FACTORS.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PERR-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL:	U59496:	AAC52730.1;	-	-	
DR	EMBL:	AF003695:	AAC53455.1;	-	-	
DR	EMBL:	Y13685:	CAA70306.1;	JOINED.		
DR	EMBL:	Y09085:	CAA70305.1;	-	-	
DR	EMBL:	AF004155:	AAC53461.1;	-	-	
DR	EMBL:	AF004141:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004142:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004143:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004144:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004145:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004146:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004147:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004148:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004149:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004150:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004151:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004152:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004153:	AAC53461.1;	JOINED.		
DR	EMBL:	AF004154:	AAC53461.1;	JOINED.		
DR	EMBL:	X95580:	CAA64833.1;	-	-	
DR	EMBL:	X95002:	CAA64458.1;	-	-	
DR	MGI:	106918:	H4fla.			
DR	InterPro:	IPR003015:	HLH_Myc.			
DR	InterPro:	IPR001092:	HLH_dim.			
DR	InterPro:	IPR001321:	HypoxindflA.			
DR	InterPro:	IPR001610:	PAC.			
DR	InterPro:	IPR000014:	PAS.			
DR	Pfam:	PF00785:	PAC; 1.			
DR	Pfam:	PF00983:	PAS; 2.			
DR	PRINTS:	PRO1080:	HYPOXIAIFLA.			
DR	SMART:	SMO0353:	HLH; 1.			
DR	SMART:	SMO0086:	PAC; 1.			
DR	SMART:	SMO0091:	PAS; 2.			
DR	PROSITE:	PS00038:	HELIX_LOOP_HELIX; 1.			
DR	PROSITE:	PS50112:	PAS; 2.			
KW	Repeat:	DNA-binding; Nuclear	Activator; Phosphorylation.			
FT	DNA_BIND	17	30			
FT	DOMAIN	31	71			
FT	DOMAIN	80	155			
FT	DOMAIN	228	298			
FT	DOMAIN	302	345			
FT	CONFLICT	31	31			
FT	CONFLICT	128	128			
FT	CONFLICT	351	351			
FT	CONFLICT	511	511			
FT	CONFLICT	686	686			
FT	CONFLICT	785	785			
Q	SEQUENCE	822 AA;	91858 MW;			

Query Match	30.3%;	Score 495;	DB 1;	Length 822;
Best Local Similarity	44.1%;	Pred. NO. 2.3e-35;		
Matches 112;	Conservative 29;	Mismatches 61;	Indels 52;	Gaps 5;

```
OY      ORVSNITELREKRSDDAARRSROETELYOLATTPFARGVSANHLKASIMRLTISYL 65
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      9 EKKMSSSRREKRSDAAARSPRKSSEVEFYLAQLOLPLPHNVSSHLDKASVMRLTISYL 68
OY      66 MHRLCAGGCKGRATGR-----LLPBGPGGFRHGTTHRRGRHGLPVKGC 108
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      69 VRKLDEGGDISEDEMRAQMDCFLYKALDGFVWLTLDG----- 107
OY      109 QAPGPQSVDCSSSLHTNPFG--TNFSLELTIGHSIFDTFHPCQEELDALFPRLSK 167
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      108 -----DMVYISDWNVKRMGLTGPF--ELAGISVYDFTHHPCDHHEMRKELTHRNPPVR 156
OY      168 KLEAPFERHESLRMKSTUSRGFTLNLIKAATWVLHCSGHMRAYPKPAOTSPPASPRSE 227
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 KGKELTNGRFFLMKCTIUSRGRTNMIKSATWVKVLCGTGHIVHYDNINSPOCG--YKK 214
OY      228 PLOCIVLICEAIP 241
        |||:||||||| ||
Db      215 PMTCVLVICEPIP 228
```

RESULT	2
HIFA_HUMAN	
ID	HIFA_HUMAN
AC	Q16665
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein) (Member of PAS protein 1) (MOP1) (HIF1 alpha).
GN	HIF1A
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxId:9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 166-170; 259-289 AND 771-781.
RX	MEDLINE=95296340; PubMed=7539918;
RA	Wang G.L., Jiang B.-H., Rue E.A., Semenza G.L.;
RT	"Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension.";
RL	Proc. Natl. Acad. Sci. U.S.A. 92:5510-5514(1995).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Hepatoma;
RX	MEDLINE=97236817; PubMed=9079689;
RA	Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RT	Przy-Grant M., Perdev G.H., Bradfield C.A.;
RL	"Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway.";
RL	J. Biol. Chem. 272:8581-8593(1997).
RN	[3]
RP	SEQUENCE FROM N.A.
RA	Rupert J.L., Hochachka P.W.;
RL	"HIF1 sequence in the Quechua, a high altitude population.";
RL	Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC	-1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.
CC	SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC	BHLH PROTEIN. HETEROODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.
CC	THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CANNOT
CC	DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90.
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. HIGHEST LEVELS IN
CC	KIDNEY AND HEART.
CC	-1- INDUCTION: UNDER REDUCED OXYGEN TENSION.
CC	-1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY



```

CC      RESIDE WITHIN THE C-TERMINAL PART.
CC      -1- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC      -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC      TRANSCRIPTION FACTORS.
CC      -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage is for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      -----
DR      EMBL; U22431; AAC50152.1; -.
DR      EMBL; U29165; AAC51210.1; -.
DR      EMBL; AF207602; AAF20139.1; -.
DR      EMBL; AF207602; AAF20140.1; -.
DR      EMBL; AF208487; AAF20149.1; -.
DR      TRANSFAC; T01610; -.
DR      MIM; 603348; -.
DR      InterPro; IPR003015; HLH_MYC.
DR      InterPro; IPR001092; HLH_dim.
DR      InterPro; IPR001321; Hypoxindf1a.
DR      InterPro; IPR001610; PAC.
DR      InterPro; IPR000014; PAS.
DR      Pfam; PF00785; PAS; 1.
DR      Pfam; PF00989; PAS; 2.
DR      PRINTS; PR01080; HYPOXIAF1A.
DR      SMART; SM00353; HLH; 1.
DR      SMART; SM00086; PAC; 1.
DR      SMART; SM00091; PAS; 2.
DR      PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR      PROSITE; PS01112; PAS; 2.
KW      Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW      Activator; Phosphorylation.
FT      DNA_BIND 17..30 BASIC DOMAIN.
FT      DOMAIN 31..71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT      DOMAIN 85..158 PAS 1.
FT      DOMAIN 228..298 PAS 2.
FT      DOMAIN 302..345 PAC.
FT      DOMAIN 615..621 POLY-THR.
SQ      SEQUENCE 826 AA; 92670 MM; ABDAF7DAA135BE2D CRC64;
Query Match 30.3%; Score 495; DB 1; Length 826;
Best Local Similarity 46.2%; Pred. NO. 2.3e-35;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
QY 11 NTELRKESRPAARSRSROETEVLTQALHTLPFARGVSAAHLIDKASIMLTSTSYLMMHLC 70
DB 14 SSKRRKESRPAARSRSRSRSESEVPELHQLPLPHNVSSHLIDKASVYMRLLTSLYLVRKLL 73
QY 71 AAGGKRGRAATGRLLPEBGGFRHGRTHGRGRLGPVGKCOA-----PGPOSDVLC 120
DB 74 DAG-----DLQIEDDMKQNMNCFYALKALDGFVNLTD 105
QY 121 SSLLIH-----NPTPG-TNFSLELIGHSIFDFIHFPCDDEIADALTTPRNLSKKKLEAPT 174
DB 106 DGDIMATYSIDNVNKKYMGLOF--ELTGHSHVDFPTPCDHEEMKEMLTTHNGILVKKKEONT 163
QY 175 ERHFSLRKSTLTSRGRTLNLKATWVKVLRHSGHARAKKPPAQTSPASSPRESEPLQCLY 234
DB 164 QRSFFLRKKTLLTSRGRTLNKTSATWVKVLRHSGHARAKKPPAQTSPASSPRESEPLQCLY 221
QY 235 LICEAIP 241
DB 222 LICEIP 228

```

ID	PAST1_MOUSE	STANDARD:	PRT:	874 AA.
DT	PF74781:	008787:	055046:	
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	01-MAR-2002	(Rel. 41, Last annotation update)		
DE	Endothelial PAS domain protein 1 (EPAS1), a transcription factor			
DE	(HLHF) (HIF-related factor) (HRF).			
GN	EPAS1.			
OC	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RC	MEDLINE=97152468; PubMed=9000051;			
RX	MEDLINE=97272213; PubMed=9119379;			
RA	Ema W., Iaya S., Yokotani N., Sogawa K., Matsuda Y.,			
RA	Fuji-I., Kuriyama Y.;			
RT	"A novel bHLH-PAS factor with close sequence similarity to hypoxia-			
RT	inducible factor 1alpha regulates the VEGF expression and its			
RT	potentially involved in lung and vascular development.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Brain capillary;			
RC	MEDLINE=97321546; PubMed=9178256;			
RX	Flamme I., Froehlich T., von Reutern M., Koppel A., Damett A.,			
RA	Risau W.;			
RA	"HRF, a putative basic helix-loop-helix-PAS-domain transcription			
RT	factor is closely related to hypoxia-inducible factor-1 alpha and			
RT	developmentally expressed in blood vessels.";			
RL	Mech. Dev. 63:51-60(1997).			
CC	-1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN			
CC	REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE			
CC	ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR			
CC	(VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF			
CC	BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE			
CC	IN THE FORMATION OF THE ENDOTHELIUM GIVING RISE TO THE BLOOD BRAIN			
CC	BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE EXPRESSION.			
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER			
CC	BHLH PROTEIN. HETERODIMER WITH THE ARNT PROTEIN.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS			
CC	IN LUNG, FOLLOWED BY HEART, KIDNEY, BRAIN AND LIVER. PREDOMINANTLY			
CC	EXPRESSED IN ENDOTHELIAL CELLS. ALSO FOUND IN SMOOTH MUSCLE CELLS			
CC	OF THE UTERUS, NEURONS, AND BROWN ADIPOSE TISSUE. HIGH EXPRESSION			
CC	IN EMBRYONIC CHOROID PLEXUS AND KIDNEY GLOMERULI.			
CC	-1- DEVELOPMENTAL STAGE: IN DAY 11 EMBRYO, EXPRESSION IS ALMOST			
CC	EXCLUSIVELY SEEN IN ENDOTHELIAL CELLS OF THE INTERSEGMENTAL BLOOD			
CC	VESSELS SEPARATING THE SOMITES, THE ATRIAL AND VENTRICULAR			
CC	CHAMBERS OF THE HEART, AND THE DORSAL AORTA. HIGH EXPRESSION ALSO			
CC	OCCURS IN EXTRA-EMBRYONIC MEMBRANES. IN THE DEVELOPING BRAIN OF			
CC	DAY 13 EMBRYO, ENDOTHELIAL CELLS OF THE HIGHLY VASCULARIZED			
CC	CHOROID PLEXUS CONTAIN HIGH LEVELS OF EPAS1.			
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF			
CC	TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.			
CC	-1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-slb.ch/announce/">http://www.isb-slb.ch/announce/</a>			





```
RESULT 6
NPAL_MOUSE STANDARD: PRT; 594 AA.
AC P97459;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronal PAS domain protein 1 (Neuronal PAS1).
GN NPAS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97165088; PubMed=9012850;
RA Zhou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,
RA Shelton J., Richardson J., Russell D.W., McKnight S.L.;
RT "Molecular characterization of two mammalian bHLH-PAS domain proteins
RT selectively expressed in the central nervous system.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).
CC -1 SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1 TISSUE SPECIFICITY: IN BRAIN, EXCLUSIVELY NEURONAL. ALSO FOUND IN
CC SPINAL CORD.
CC -1 DEVELOPMENTAL STAGE: FIRST DETECTED BETWEEN EMBRYONIC DAY 15 AND
CC DAY 16.
CC -1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1 SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1 SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U77967; AAB47247.1; -.
DR MGD: MGI:109205; Npas1.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00036; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS00112; PAS; 1.
DR Repeat: DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 46 58 BASIC DOMAIN.
FT DOMAIN 59 99 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 135 205 PAS 1.
FT DOMAIN 294 360 PAS 2.
FT DOMAIN 366 409 PAC.
FT DOMAIN 212 223 POLY-SER.
SQ SEQUENCE 594 AA; 63736 MW; D6477696DF69A4B3 CRC64;
```

Query Match 17.9%; Score 293.5; DB 1; Length 594;  
Best Local Similarity 33.1%; Pred. NO. 5.1e-18;  
Matches 101; Conservative 32; Mismatches 63; Indels 109; Gaps 15;

```
OY 15 RKESRDAAARSRQSEFVLYOLAHHTLPFARGVSADLKASIMRLTISLYLMMHRLCAAGG 74
DB 46 RKESRNAARWRKGNLEFFELAKLPPLPGAISQDLKASIVRLSVTLRLRFPALGA 105
```

```
OY 75 KRG--RATGRLPEGCGFRHGRHGRGLPYGKQQAQP---QSYDLCSLSLHNPT 129
DB 106 PPMGLRAVG-----PPAGLAPG--RRGPVALYSEVEFHGLHILQSLD----- 147
OY 130 PGTNFG-----ELIGHSIFDHPDCDELODAL----- 159
DB 148 -GRVFALNDGKFLYISEVTSYTLGLSQVELTSSVFDYTHHPDHSVELEQLRAASIG 206
OY 160 -TPRPNLSKK-----KLEA-PT-----ERHPSLRMKSTLTSGRFTLM 195
DB 207 PPTP-PSVSSSSSSSSSLVDPTPEIASEASPARFAQGSFVRMKSTLTKRG--LNV 263
OY 196 KAATKWLKHCGHMRAVK-----PPAQTSPPASPSSEPLQ-----CLV 224
DB 264 KASGKYVHWYGRRLRARALGLVALGHTLPRA-----PLAELPLGHMIVFRLSLGLTI 316
OY 235 LICEA 239
DB 317 LACES 321
RESULT 7
STRA_DROME STANDARD: PRT; 1507 AA.
AC Q24167; Q9VAA5;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Similar protein.
GN STRA OR CG7951.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96269413; PubMed=8662312;
RA Nambu J.R., Chen W., Hu S., Crews S.T.;
RT "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein
RT related to human hypoxia-inducible factor 1 alpha and Drosophila
RT single-minded.";
RL Gene 172:249-254 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Andrews-Plamkoch C., Baldwin D.,
RA Abail J.F., Aghayani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Benson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ipegyam C.,
RA Jatalil M., Kalush F., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sviderskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Welschenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Zheng J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.  
 CC -1- SUBUNIT: EFFICIENT DNA-BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN THE EMBRYO.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U43090; AAC47303.1; -.  
 DR EMBL: AE003772; AAF57008.2; -.  
 DR FlyBase: Fgn0015542; HLM\_Myc.  
 DR InterPro: IPR003015; HLM\_Myc.  
 DR InterPro: IPR001092; HLM\_dlm.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAS; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00353; HLM; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;  
 KW Activator; Coiled coil.  
 KW -----  
 FT DNA\_BIND 72 85  
 FT DOMAIN 167 240  
 FT DOMAIN 307 377  
 FT DOMAIN 381 422  
 FT DOMAIN 577 587  
 FT DOMAIN 880 908  
 FT DOMAIN 982 1054  
 FT DOMAIN 1110 1162  
 FT DOMAIN 126 39  
 FT DOMAIN 718 725  
 FT DOMAIN 759 763  
 FT DOMAIN 767 776  
 FT DOMAIN 907 918  
 FT DOMAIN 945 948  
 FT DOMAIN 990 998  
 FT DOMAIN 1020 1038  
 FT DOMAIN 1113 1126  
 FT DOMAIN 1146 1162  
 FT DOMAIN 1205 1208  
 FT DOMAIN 1277 1284  
 FT DOMAIN 1298 1301  
 FT DOMAIN 1301 1301  
 FT CONFLICT 38 38  
 FT CONFLICT 345 345  
 FT CONFLICT 492 492  
 FT CONFLICT 588 588  
 T -> I (IN REF. 1).  
 S -> A (IN REF. 1).  
 S -> L (IN REF. 1).  
 A -> V (IN REF. 1).  
 T -> I (IN REF. 1).

FT CONFLICT 709 709 T -> K (IN REF. 1).  
 FT CONFLICT 776 776 Q -> Q (IN REF. 1).  
 FT CONFLICT 895 895 Q -> Q (IN REF. 1).  
 FT CONFLICT 902 902 Q -> S (IN REF. 1).  
 FT CONFLICT 982 982 A -> T (IN REF. 1).  
 FT CONFLICT 1125 1126 MISSING (IN REF. 1).  
 FT CONFLICT 1154 1157 MISSING (IN REF. 1).  
 FT CONFLICT 1444 1444 F -> L (IN REF. 1).  
 FT CONFLICT 1447 1447 G -> C (IN REF. 1).  
 FT CONFLICT 1451 1451 G -> N (IN REF. 1).  
 FT CONFLICT 1494 1494 D -> G (IN REF. 1).  
 SQ SEQUENCE 1507 AA; 165824 MW; 4102939C8FBF00C6 CRC64;  
 Query Match 17.2%; Score 281; DB 1; Length 1507;  
 Best Local Similarity 30.2%; Pred. No. 1.9e-16;  
 Matches 71; Conservative 47; Mismatches 71; Indels 46; Gaps 5;  
 QY 9 RSNTELKRSRDAARSRSETEVLYQLAHTLPFARGVSAHLDPKASIMRLTSLYLMHR 68  
 DB 67 RRNNEKRSRDAARSRSETEVLYQLAHTLPFARGVSAHLDPKASIMRLTSLYLMHR 126  
 QY 69 L-----CAAGKRGATGRLLPAGPGFRHGRHGRHGLPVKCOQAPGQSVLD 119  
 DB 127 MLDVPSLRDNDIKQDIETAE-----DQEVKPKLEVTEMDLNGAEAREL 174  
 QY 120 CSSSL-----INPPTGTFNS-----LELGHSFTDFIPPCDDELADALTPR 162  
 DB 175 LKQTMDFLVLSHEGDITVSENVVEYLGITKIDTIGQOQWESHOCHAEIEALSLK 234  
 QY 163 PNLKKLEAPTE-----RHFLMKSTLTSRGRTLNKAATWVLCOSGM 209  
 DB 235 RELAQKVDPEQNSGVSTHHRDLFVRLKCTLISRGSINIKSAYIVIHITGHL 289  
 RESULT 8  
 SIML\_MOUSE STANDARD; PRT; 765 AA.  
 ID SIML\_MOUSE  
 AC 061045; P70183;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-minded homolog 1 (MSIM1).  
 GN SIM1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER;  
 RX MEDLINE=97020303; PubMed=8812055;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RT "Expression patterns of two murine homologs of *Drosophila*  
 RT single-minded suggest possible roles in embryonic patterning and in  
 RT the pathogenesis of Down syndrome.";  
 RT Mol. Cell. Neurosci. 7:1-16(1996).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=97029422; PubMed=8875433;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RL Mol. Cell. Neurosci. 7:519-519(1996).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RX MEDLINE=97343329; PubMed=9199934;  
 RA Fan C.-M.;  
 RL Unpublished results, cited by:  
 RL Christ R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,  
 RL Wang Y., Shimizu N., Antonarakis S.E.;

RL Genome Res. 7:615-624(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J;  
 RX MEDLINE-96413339; PubMed-8927054;  
 RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,  
 RA Saijoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;  
 RT "Two new members of the murine sim gene family are transcriptional  
 RT repressors and show different expression patterns during mouse  
 RT embryogenesis";  
 RL Mol. Cell. Biol. 16:5865-5875(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RX Hosoya T.;  
 RN Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.  
 RP SUBUNIT.  
 RX MEDLINE-97172525; PubMed-9020169;  
 RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;  
 RT "Two murine homologs of the Drosophila single-minded protein that  
 RT interact with the mouse aryl hydrocarbon receptor nuclear  
 RT translocator protein";  
 RL J. Biol. Chem. 272:4451-4457(1997).  
 CC - FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC - TISSUE SPECIFICITY: DETECTED IN LUNG, SKELETAL MUSCLE AND KIDNEY.  
 CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING  
 CC KIDNEY, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING  
 CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.  
 CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U40575; AAA91201.1; ALT\_SEQ.  
 DR EMBL: D79209; BAA11467.1; -.  
 DR EMBL: AB013491; BAA28270.1; -.  
 DR EMBL: AB013484; BAA28270.1; JOINED.  
 DR EMBL: AB013485; BAA28270.1; JOINED.  
 DR EMBL: AB013486; BAA28270.1; JOINED.  
 DR EMBL: AB013487; BAA28270.1; JOINED.  
 DR EMBL: AB013488; BAA28270.1; JOINED.  
 DR EMBL: AB013489; BAA28270.1; JOINED.  
 DR EMBL: AB013490; BAA28270.1; JOINED.  
 DR MGD: MGI:96306; Sim1.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dlm.  
 DR InterPro: IPR001067; Nucleinslocator.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR PRINTS: PM00785; NCTRNSLOCATR.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT DNA\_BIND 1 13 BASIC DOMAIN  
 FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 77 147 PAS 1.  
 FT DOMAIN 218 288 PAS 2.  
 FT CONFLICT 133 133 H -> L (IN REF. 1).  
 FT CONFLICT 176 176 MISSING (IN REF. 1).  
 FT CONFLICT 322 322 P -> R (IN REF. 1).  
 FT CONFLICT 480 480 A -> P (IN REF. 1).  
 FT CONFLICT 537 537 D -> S (IN REF. 1).  
 SO SEQUENCE 765 AA; 85540 MW; B1AF7DA8578CD17 CRC64;  
 Query Match 16.4%; Score 268; DB 1; Length 765;  
 Best Local Similarity 36.2%; Pred. No. 1,1e-15;  
 Matches 77; Conservative 28; Mismatches 86; Indels 22; Gaps 6;  
 QY 16 KEKSPAAARSRSQETEVYLAHTLPFAGVSAHLDDKASIMLTJTYLMBHRLCAAG-G 74  
 DB 2 KEKSKNAARTRREKENSEFEELAKLLPLPSAITSQDKASTILNTSYLKMAYVPEGGIG 61  
 QY 75 KRGKATGRLPEGPGRHGT-RRGRHGL-----PVKCOQAPGQSYDLCSSLIHNP 128  
 DB 62 EAMGHTSRSPDLNVGRELGSHLQTLDFIFVAVPDGKIMYISFASVHLGTS----- 115  
 QY 129 TPGTNSLELGHSIFDFIHPQDELQDALTP-RNLSKKLEAFERHFSLRMKSTLT 187  
 DB 116 -----QVELTGNISYIEYHPADHDMTAVLTAAHQPHSHFVQVEIERSEFSLKMCVLA 169  
 QY 188 SRGRTLNKAATKWLKVGSGHMRAYKPPAQTSP 220  
 DB 170 K--RNAGLICGKIVHCISGLKIRQYSLDMSP 200  
 RESULT 9  
 SIM1\_HUMAN  
 ID SIM1\_HUMAN STANDARD: PRT; 766 AA.  
 AC P81133;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-minded homolog 1.  
 GN SIM1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97343329; PubMed-9199934;  
 RA Christ R., Scott H.S., Chen H., Kudoh J., Rossler C.,  
 RA Minoshima S., Wang Y., Shimizu N., Antonarakis S.E.;  
 RT "Cloning of two human homologs of the Drosophila single-minded gene  
 RT SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome  
 RT chromosomal region";  
 RL Genome Res. 7:615-624(1997).  
 CC - FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U70212; AAB62395.1; -.  
 DR MIM: 603128; -.  
 DR InterPro: IPR003015; HLH\_Myc.





```
FT CONFLICT 382 430 RTNPPPOOYSSFOQMDKLECSOYGNMTSPPTNAVAPPEQ
FT CONFLICT 501 501 LHSEASDL -> KNOPISPPAPSKWTWMSAARWETGELVPP
FT CONFLICT 541 541 OMLPRNSSIOCKPAT (IN REF. 3).
FT CONFLICT 561 585 P -> T (IN REF. 3).
FT CONFLICT 561 585 P -> R (IN REF. 3).
FT CONFLICT 590 591 APPROASRDARLALARPPECCARP -> VLARRPGHAROM
FT CONFLICT 638 638 MES (IN REF. 3).
FT CONFLICT 641 657 OA -> HG (IN REF. 3).
FT CONFLICT 641 657 A -> R (IN REF. 2).
FT CONFLICT 641 657 APPHYGASVITNGR -> PPTALPGRLHHOROVTRQ
FT CONFLICT 641 657 SRSAGRPSGSHHQAHSBSGSPTRKVOQMEGRYVARKLY
FT CONFLICT 641 657 TETTLLEKSLRDLDFTRCWCGLGSGWEMGOTAHSCPR
FT CONFLICT 641 657 AGFGAHEHPVNL (IN REF. 3).
FT CONFLICT 641 657 C7904CD24COABBAF CRC64;
SQ SEQUENCE 657 AA; 72512 MW; C7904CD24COABBAF CRC64;

Query Match 16.28; Score 265; DB 1; Length 657;
Best Local Similarity 35.78; Pred. No. 1.7e-15;
Matches 74; Conservative 26; Mismatches 77; Indels 30; Gaps 6;

QY 16 KEKRDAAARSRQSEYLYOLATLPFARGVSAHLKASIMRLTISYLRHRLCAAGK 75
DB 2 KEKKNNAKTRREKENEFYELAKLPLPSATISQDLKASITRLTISYLRHRLCAAGK 59
QY 76 RGRATGRLPEGPGFRRHGRHGRHGLPV-----GKCOQAPGPOVDLCSSSL 124
DB 60 LGDAMGQ--PSRTPGLDSVAKELGSHLQTLDFGVVVASDGIIMYISFSAVHLGLS-- 115
QY 125 IHNTPGTNSLELIGHSIFPIHPCOEELQDALTPRPNLSKKL-EAPTERHFSLRMK 183
DB 116 -----QVELTGNISYIEYHPSDHEMTAVLTAPHLHLLDEYELERSFFLEMK 165
QY 184 STLSRGRTLNKAATWKLCHSGHMR 210
DB 166 CVLAK--RMAGLTGSGYKVIHCSTYK 190

RESULT 11
SIM2_HUMAN STANDARD; PRT; 667 AA.
ID SIM2_HUMAN STANDARD; PRT; 667 AA.
AC 014190: Q15470; Q15471; Q15473; Q15473; Q15473; Q15473;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 2.
GN SIM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97343329; PubMed=9199934;
RA Chraist R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,
RA Wang Y., Shimizu N., Antonarakis S.E.;
RT "Cloning of two human homologs of the Drosophila single-minded gene
RT SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome
RT chromosomal region."
RT Genome Res. 7:615-624(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Shintani A., Asakawa S., Shimizu N.;
RT "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2."
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Shintani A., Asakawa S., Shimizu N.;
RT "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2."
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-247 FROM N.A.
RX MEDLINE=96016135; PubMed=7568099;
RA Oseogawa K., Okano S., Soeda E.;
RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-58; 87-152; 183-247 AND 249-283 FROM N.A.
RX MEDLINE=95375794; PubMed=7647800;
RA Chen H., Chraist R., Rossier C., Gos A., Antonarakis S.E., Kudoh J.,
RA Yamaki A., Shindoh N., Maeda H., Minoshima S., Shimizu N.;
RT "Single-minded and Down syndrome?";
RT Nat. Genet. 10:9-10(1995).
RN [6]
RP SEQUENCE OF 87-116 FROM N.A.
RX MEDLINE=96016135; PubMed=7568099;
RA Dahmane N., Charbon G., Lopes C., Yaspo M.-L., Maunouy C.,
RA Decorte L., Smet P.M., Bloch B., Delabar J.M.;
RT "Down syndrome-critical region contains a gene homologous to
RT Drosophila sim expressed during rat and human central nervous system
RT development.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:9191-9195(1995).
RN [7]
RP SEQUENCE OF 154-181 FROM N.A.
RX MEDLINE=96299750; PubMed=8661114;
RA Yamaki A., Noda S., Kudoh J., Shindoh N., Maeda H., Minoshima S.,
RA Yamaki K., Shimizu Y., Shimizu N.;
RT "The mammalian single-minded (SIM) gene: mouse cDNA structure and
RT dienecephalic expression indicate a candidate gene for Down
RT syndrome.";
RT Genomics 35:136-143(1996).
RN [8]
RP FUNCTION: TRANSCRIPTION FACTOR THAT MAY BE A MASTER GENE OF CNS
DEVELOPMENT IN COOPERATION WITH ARNT. IT MAY HAVE PLEIOTROPIC
EFFECTS IN THE TISSUES EXPRESSED DURING DEVELOPMENT.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
BHLH PROTEIN. HETERODIMER OF SIM2 AND ARNT.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIM2 (SHOWN HERE) AND SIM2S, ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U80457; AAB62397.1; -
DR EMBL: U80457; AAB62397.1; -
DR EMBL: AB003185; BAA21489.1; -
DR EMBL: AB003185; BAA21490.1; -
DR EMBL: D85922; BAA12919.1; -
DR EMBL: D44444; BAA07906.1; -
DR EMBL: D44445; BAA07907.1; -
DR EMBL: D44446; BAA07908.1; -
DR EMBL: D44447; BAA07909.1; -
DR EMBL: D44448; BAA07910.1; -
DR EMBL: X84790; CAA59261.1; -
DR EMBL: D70838; BAA1108.1; -
DR EMBL: AP000697; BAA89433.1; -
DR EMBL: AP001726; -; NOT_ANNOTATED_CDS.
DR MIM: 600892; -;
```





CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U33427; AAA96257.1; ALT\_INIT.  
DR EMBL: U42699; AAA96754.1; ALT\_FRAME.  
DR EMBL: AE003468; AAR47386.1; -.  
DR EMBL: AE003468; AAR47386.1; trh.  
DR InterPro: IPR003015; HLH\_MYC.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00785; PAC; 1.  
DR Pfam: PF00989; PAS; 2.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
DR PROSITE: PS0112; PAS; 2.  
KW Developmental protein; Nuclear protein; Transcription regulation;  
KW Repeat; DNA-binding; Alternative splicing.  
FT DNA\_BIND 86 99 BASIC DOMAIN.  
FT DOMAIN 100 140 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 174 244 PAS 1.  
FT DOMAIN 391 461 PAS 2.  
FT DOMAIN 465 508 PAC.  
FT DOMAIN 629 636 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 154 157 POLY-SER.  
FT DOMAIN 244 250 POLY-GLY.  
FT DOMAIN 251 254 POLY-SER.  
FT DOMAIN 255 264 POLY-GLN.  
FT DOMAIN 722 728 POLY-SER.  
FT VARSPLIC 281 286 MISSING (IN ISOFORM 2).  
FT VARSPLIC 328 356 MISSING (IN ISOFORM 3).  
FT CONFLICT 78 78 G -> A (IN REF. 1).  
FT CONFLICT 250 250 G -> GG (IN REF. 2).  
FT CONFLICT 703 703 A -> T (IN REF. 1).  
FT CONFLICT 708 708 A -> P (IN REF. 1).  
FT CONFLICT 829 829 A -> V (IN REF. 1).  
SQ SEQUENCE 958 AA; 102241 MW; 8F9CF758E1370541 CRC64;  
  
Query Match 15.3%; Score 251; DB 1; Length 958;  
Best Local Similarity 27.2%; Pred. No. 4.5e-14;  
Matches 89; Conservative 41; Mismatches 91; Indels 106; Gaps 10;

Db 371 SHSRKSPPLGMAVALATAPPPSVHE 397  
RESULT 13  
SIM\_DROME STANDARD; PRT; 673 AA.  
AC P05709; Q9VPE23;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Single-minded protein.  
GN SIM OR CG7771.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE OF 19-673 FROM N.A.  
RX MEDLINE=88151023; PubMed=3345560;  
RA Crews S.T., Thomas J.B., Goodman C.S.;  
RT "The Drosophila single-minded gene encodes a nuclear protein with  
RT sequence similarity to the per gene product.";  
RL Cell 52:143-151(1988).  
RN [2]  
RP SEQUENCE OF 1-18 FROM N.A., AND SIMILARITY TO HLH PROTEINS.  
RX MEDLINE=92103681; PubMed=1760843;  
RA Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;  
RT "The Drosophila single-minded gene encodes a helix-loop-helix protein  
RT that acts as a master regulator of CNS midline development.";  
RL Cell 67:1157-1167(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99054545; PubMed=9840810;  
RA Kasai Y., Stehl S., Crews S.;  
RT "Specification of the Drosophila CNS midline cell lineage: direct  
RT control of single-minded transcription by dorsal/ventral patterning  
RT genes.";  
RL gene Expr. 7:1171-1189(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BERKELEY;  
FC MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Henderson S.N.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abriell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Fodor C.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garb N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER  
 CC DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN  
 CC THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING  
 CC RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M19020; AAA28900.1; -  
 DR EMBL: AF071934; AAC64519.1; ALT\_SEQ.  
 DR EMBL: AE003698; AAF54902.1; ALT\_SEQ.  
 DR PIR: A29945; A29945.  
 DR TRANSFAC: T00750; -  
 DR FlyBase: FBgn000466; sim.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dlm.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC. 1.  
 DR Pfam: PF00989; PAS. 2.  
 DR SMART: SM00353; HLH. 1.  
 DR SMART: SM0086; PAC. 1.  
 DR SMART: SM0091; PAS. 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX. 1.  
 DR PROSITE: PS00112; PAS. 2.  
 KM Developmental protein: Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT Dna\_BIND 1 13  
 FT DOMAIN 14 54 BASIC DOMAIN.  
 FT DOMAIN 76 148 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 242 312 PAS 1.  
 FT DOMAIN 382 422 PAS 2.  
 FT DOMAIN 423 426 14 X 3 AA REPEATS OF A-A-Q.  
 FT DOMAIN 484 488 POLY-VAL.  
 FT DOMAIN 489 492 POLY-SER.  
 FT DOMAIN 496 499 POLY-ASN.  
 FT DOMAIN 501 505 POLY-HIS.  
 FT DOMAIN 529 533 POLY-GLN.  
 FT DOMAIN 529 533 POLY-SER.  
 FT DOMAIN 573 576 POLY-ASN.  
 FT DOMAIN 604 607 POLY-SER.  
 FT DOMAIN 649 659 GLN/HIS-RICH.  
 FT DOMAIN 127 127 I -> Y (IN REF. 3).  
 FT CONFLICT 401 409 MISSING (IN REF. 4).  
 SO SEQUENCE 673 AA; 73589 MW; 2E9FOABBA2BC0FBE CRC64;

Query Match 15.2% Score 249; DB 1; Length 673;  
 Best Local Similarity 30.1% Pred. No. 4.4e-14;  
 Matches 83; Conservative 39; Mismatches 84; Indels 70; Gaps 10;

OY 16 KEKRDARSRSQETVLYQLAHTLPAPRGVSAHLDAKASIMRLTISLRHRLCAAGK 75  
 DB 2 KEKSNNAIRREKEMFECLAKLLPAPAITOLDKASVIRLTSTYLRKROVPPDG-- 59

OY 76 RGRATGRLLPEGPCGFRHGRHRR-GRHGL-----PVKCGQAPGPOSYDLCSSS 123  
 DB 60 LGEMAG-----SSPAMQGRATIKELGSHLQTLDFIFVAPADGKIMYSERASVHLGLS- 114  
 OY 124 LIHNPPTGNTSFLDIGHISFDFHPCQDEELQALTPRPMLSKRL----- 170  
 DB 115 -----QVELGNSIFEYIHNYQDEMNALISLPHINQHPLAQTHTPIGSPNGV 163  
 OY 171 EAPF-----ERHSLRKSKTLTSGRLNLNKAATWKLHSCGHHRAKPPAO 217  
 DB 164 QHPSAYDHDKRSHTIEKTFELRMKCVLAR--RAAGLTTSGFKVHSGYKARIYDR 221  
 OY 218 TSPAGSPSEPPLQCLVLC-----EAPLPDPH 246  
 DB 222 GDGGS-----LIQNLGLVANGHSILPSSAITFEIKH 252  
 RESULT 14  
 CYCL\_DROME STANDARD: PRT: 413 AA.  
 AC 061734; 076344; 09YV44;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cycle protein (Brain and muscle ARNT-like 1) (BMAL1) (MOP3).  
 GN CYC OR CG8727.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID=7227;  
 RX MEDLINE-98292178; PubMed-9630224;  
 RT "CYCLE is a second bHLH-PAS clock protein essential for circadian  
 RT rhythmicity and transcription of *Drosophila* period and timeless.";  
 RL Cell 93:805-814(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain, and Muscle;  
 RX MEDLINE-98279147; PubMed-9616122;  
 RA Darlington T.K., Mager-Smith K., Ceriani M.F., Stekakis D., Gekakis N.,  
 RA Steeves T.D.L., Weitz C.J., Takahashi J.S., Kay S.A.;  
 RT "Closing the circadian loop: CLOCK-induced transcription of its own  
 RT inhibitors per and tim.";  
 RL Science 280:1599-1603(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abciril J.F., Abghayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Boucek J., Brokstein P., Brotlier P.,  
 RA Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunhov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,





...

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2002, 07:34:02 ; Search time 62.38 Seconds  
(without alignments)  
851.386 Million cell updates/sec

Title: US-09-896-791B-3  
Perfect score: 1636  
Sequence: 1 MALGIGRVSRNTLEKREKR.....TESSLPSVWLALNRKNCPC 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855.5	52.3	630	11 Q9QX54	Q9qX54 mus musculu
2	855.5	52.3	662	11 Q9Z215	Q9Z215 mus musculu
3	831	50.8	632	4 Q96K34	Q96K34 homo sapien
4	820.5	50.2	662	11 Q9JHS2	Q9JHS2 rattus norv
5	802	49.0	667	4 Q9Y2N7	Q9Y2N7 homo sapien
6	573	35.0	648	4 Q9HAI2	Q9hAI2 homo sapien
7	518.5	31.7	199	4 Q9HAM3	Q9ham3 homo sapien
8	503	30.7	811	13 Q9YIB9	Q9yIB9 gallus gall
9	499	30.5	823	6 Q9XTAS	Q9xtas bos taurus
10	497	30.4	825	11 Q35800	Q35800 rattus norv
11	495	30.3	258	11 Q9CYA8	Q9cyA8 mus musculu
12	495	30.3	735	4 Q96PT9	Q96pt9 homo sapien
13	495	30.3	826	4 Q9UPB1	Q9upb1 homo sapien
14	491	30.0	823	11 Q9WT09	Q9wtu9 rattus norv
15	482	29.5	874	11 Q9JHS1	Q9jhs1 rattus norv
16	471	28.8	867	13 Q9W7C6	Q9w7c6 gallus gall

17	471	28.8	870	13 Q9PTB3	Q9ptb3 coturnix co
18	468	28.6	760	6 Q9XTA4	Q9xtA4 bos taurus
19	450	27.5	876	13 Q9BSW2	Q9BSW2 oncorhynch
20	426	26.0	237	4 Q9H7Z9	Q9h7Z9 homo sapien
21	322	19.7	115	4 Q95262	Q95262 homo sapien
22	312	19.1	112	4 Q9UPH7	Q9upH7 homo sapien
23	297.5	18.2	805	13 Q918A9	Q918A9 xenopus lae
24	297	18.2	590	4 Q9BYR3	Q9byR3 homo sapien
25	286.5	17.5	925	11 Q9QZ00	Q9qZ00 mus musculu
26	278.5	17.0	901	4 Q9H323	Q9h323 homo sapien
27	276.5	16.9	903	4 Q9BR81	Q9br81 homo sapien
28	266	16.3	765	11 Q70284	Q70284 mus musculu
29	265	16.2	657	11 Q35391	Q35391 mus musculu
30	257.5	15.7	849	5 Q15984	Q15984 bombyx mori
31	255	15.6	760	13 Q9PD06	Q9pd06 xenopus lae
32	238	14.5	745	13 Q9BSJ5	Q9BSJ5 brachydanto
33	169.5	10.4	719	5 Q963U8	Q963U8 caenorhabd1
34	169.5	10.4	823	5 Q45486	Q45486 caenorhabd1
35	169.5	10.4	825	5 Q9TYM0	Q9tyM0 caenorhabd1
36	166	10.1	531	11 Q9Z3L9	Q9Z3L9 rattus norv
37	158.5	9.7	575	13 Q9IAU1	Q9iaU1 brachydanto
38	157.5	9.6	199	4 Q9Y4L8	Q9y4L8 homo sapien
39	157	9.6	103	6 Q9N110	Q9n110 ovis aries
40	156.5	9.6	534	11 Q9Z3L8	Q9Z3L8 rattus norv
41	156	9.5	565	11 Q9Z4H3	Q9Z4H3 rattus norv
42	154	9.4	87	13 Q90692	Q90692 gallus gall
43	153.5	9.4	852	5 Q9NC54	Q9nc54 mya arenari
44	149	9.1	108	11 Q90Z94	Q9qZ94 cavia porce
45	142	8.7	805	13 Q9BSN3	Q9BSN3 gallus gall

#### ALIGNMENTS

RESULT 1  
ID Q9QX54 PRELIMINARY; PRT; 630 AA.  
AC Q9QX54;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOXIA-INDUCIBLE FACTOR 3 ALPHA (FRAGMENT).  
GN HIF3A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
RT "Cloning and Characterization of a Third Hypoxia Inducible Factor,  
RT HIF3-alpha.";  
RT J. Biol. Chem. 0:0-0(1999).  
RL EMBL; AF079153; AAF21782.1;  
DR EMBL; AF079140; AAF21782.1; JOINED.  
DR EMBL; AF079141; AAF21782.1; JOINED.  
DR EMBL; AF079142; AAF21782.1; JOINED.  
DR EMBL; AF079143; AAF21782.1; JOINED.  
DR EMBL; AF079144; AAF21782.1; JOINED.  
DR EMBL; AF079145; AAF21782.1; JOINED.  
DR EMBL; AF079146; AAF21782.1; JOINED.  
DR EMBL; AF079147; AAF21782.1; JOINED.  
DR EMBL; AF079148; AAF21782.1; JOINED.  
DR EMBL; AF079149; AAF21782.1; JOINED.  
DR EMBL; AF079150; AAF21782.1; JOINED.  
DR EMBL; AF079151; AAF21782.1; JOINED.  
DR EMBL; AF079152; AAF21782.1; JOINED.  
MGD: MGI:1859778; Hif3a.  
DR InterPro: IPR001092; HLH\_dlm.  
DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001067; NucleinsLocator.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.

DR pfam, PF00989; PAS; 2.  
DR PRINTS, PR00785; NCTRNSLOCATR.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM0086; PAC; 1.  
DR SMART; SM0091; PAS; 2.  
DR PROSITE, PS0038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
FT NON\_TER 630 630  
SQ SEQUENCE 630 AA; 69624 MW; 828EB2C84E6D45B6 CRC64;

Query Match	52.3%;	Score 855.5;	DB 11;	Length 630;
Best Local Similarity	68.4%;	Pred. No. 1.3e-77;		
Matches 182;	Conservative	5;	Mismatches 24;	Indels 55;
				Gaps 4;

```

Oy      9  RSNLRLKREKSRDARSRSROSETEVYLQALHTPLPFAFGVSAHLIDKASIMPLTISYLMNR  68
Db      7  RSNLRLKREKSRDARSRSROSETEVYLQALHTPLPFAFGVSAHLIDKASIMPLTISYLMNR  66
Oy      69  LCAAGKKGKRGATGRLLRPGCGEGRNHTGRKRGHGLPVGCKQQAAPGSVDLC-----  120
Db      67  LCAAGEMN-----OVEKGSGSPDLACYKALCEGF  94
Oy      121  -----SSSLIHNPPTGPNFS-LFLIGHSIFDFNPPDOEQLDQALPRRNLSSKKL  170
Db      95  VMTLTABSDMAYLSEBNVSKHLGLSLQLELIGHSIFDFNPPDOEQLDQALPRRNLSSKKL  154
Oy      171  EAPTRHPSLMAKSTLTSRGRRTLMLKAAATVWLHCSGIMRAYKRPAPOTSAGSPRSRPL  230
Db      155  EAPTRHPSLMAKSTLTSRGRRTLMLKAAATVWLHCSGIMRAYKRPAPOTSAGSPRSRPL  214
Oy      231  QCLVLCALP-----QLPFDGATL  251
Db      215  QCLVLCALPAPASLEPPLGRGAFPL  240

```

RESULT	2			
Q92215				
ID	Q92215	PRELIMINARY:	PRT:	662 AA.
AC	Q92215;			
DT	01-MAY-1999 (TrEMBLrel. 10, Created)			
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	HYPOXIA INDUCIBLE FACTOR THREE ALPHA.			
GN	HIF3A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Gu Y.Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;			
RT	"Molecular characterization and chromosomal localization of a third			
RL	alpha-class hypoxia induced factor subunit, Hif3alpha.";			
DR	EMBL; AF060194; AAC72734.1; -			
DR	MGD; MGI:1859778; Hif3a.			
DR	InterPro: IPR001092; HLH.dlm.			
DR	InterPro: IPR001015; HLH_MyC.			
DR	InterPro: IPR001067; HLH_NucR.			
DR	InterPro: IPR001610; PAC.			
DR	InterPro: IPR000014; PAS.			
DR	Pfam; PF00989; PAS; 2.			
DR	PRINTS; PR00785; NCTRSL0CATR.			
DR	SMART; SM00353; HLH; 1.			
DR	SMART; SM00086; PAC; 1.			
DR	SMART; SM00091; PAS; 2.			
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.			
DR	SEQUENCE 662 AA; 73013 MW; 58740A1B6993B3B5 CRC64;			

Query Match	52.3%;	Score 855.5;	DB 11;	Length 662;
Best Local Similarity	68.4%;	Pred. No. 1.4e-77;		
Matches 182; Conservative	5;	Mismatches 24;	Indels 55;	Gaps 4;

[illegible]

RESULT	3		
Q96K34			
ID	Q96K34	PRELIMINARY;	PRT; 632 AA.
AC	Q96K34;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)		
DE	CDNA FLJ14819 F1S, CLONE OVARC1000241, MODERATELY SIMILAR TO		
DE	HYPOXIA-INDUCIBLE FACTOR 1 ALPHA.		
OS	Homo sapiens (Human) .		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OVARIAN CARCINOMA;		
RA	Iisoga T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Mishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,		
RA	Tanai H., Yamada M., Watanabe M., Hiraoka S., Ishi S., Kawai Y.,		
RA	Sato K., Yamamoto Y., Wakamatsu A., Nakamura Y., Negahari K.,		
RT	Masuhio Y., Kanehori K.;		
RT	"NEBD human cDNA sequencing project ";		
RL	Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AK021725; BAB5324.1; ..		
SK	SEQUENCE 632 AA; 68934 MW; A19FIED3D05E7A71 CRC64;		

Query Match	50.8%;	Score 831;	DB 4;	Length 632;
Best Local Similarity	68.2%;	Pred. No. 3.8e-75;		
Matches 178;	Conservative 11;	Mismatches 30;	Indels 42;	Gaps 5.

[illegible]



```

RESULT 4
09JHS2 PRELIMINARY; PRT: 662 AA.
AC 09JHS2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPoxIA INDUCIBLE FACTOR 3 ALPHA.
GN HIF-3A.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21134367; PubMed=11237857;
RA Kletzmann T., Cornesse Y., Brechtel K., Modaresi S., Jungermann K.;
RT "Perinatal expression of the mRNA of the three hypoxia-inducible
RT factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";
RL Biochem. J. 354:531-537(2001).
DR EMBL: AJ277827; CAB96611.1; -.
DR InterPro: IPR001092; HLH dim.
DR InterPro: IPR003015; HLH MYC.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR PRINTS: PR00785; NCTRNSLOCATR.
DR SMART: SM00353; HLH. 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 662 AA; 72887 MW; AC9672E340544010 CRC64;

```

```

Query Match 50.2%; Score 820.5; DB 11; Length 662;
Best Local Similarity 68.1%; Pred. No. 4.6e-74;
Matches 177; Conservative 9; Mismatches 31; Indels 43; Gaps 5;

QY 9 RSNTELREKSRDARSRSOETEVLYQALHTLPFARGVSAHLDKASIMRLTISYLRMR 68
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 7 RSTTELREKSRDARSRSOETEVLYQALHTLPFARGVSAHLDKASIMRLTISYLRMR 66
QY 69 LCAAGKRGATGRLDPGPGFRHGTNRGRHGLPVGKCO-----QAPGPOSV 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 67 LCAAG-----EMNQYKKEGEPDADCYLKALEGVWVLTAEQDAY 106
QY 118 DLCSSTLIHNPPTGNTS-LELIGHSTFDTHPCDOELDALTPRPMLSKKLEAPTER 176
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 107 -----LSENVSKHLGSQLLELGHSTFDTHPCDOELDALTPRPMLSKKLEAPTER 160
QY 177 HFSLRMKSTLTSRGRITNLKATWKVLCGSHMRAYRPPAOTSPAGSPRSPPLQCLVLI 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 161 HFSLRMKSTLTSRGRITNLKATWKVLCGSHMRAYRPPAOTSPAGSPRSPPLQCLVLI 220
QY 237 CEALP-----QLPHDQATL 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 221 CEALPHPASLEPPLGGAFL 240

RESULT 5
09Y2N7 PRELIMINARY; PRT: 667 AA.
ID 09Y2N7;
AC 09Y2N7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE HOMOLOG OF HYPoxIA INDUCIBLE FACTOR THREE ALPHA
DE (HYPoxIA-INDUCIBLE FACTOR-3 ALPHA).
GN HIF-3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Altix C., Andreise T., Frankheim M., Amco-Keller G., Coeffield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Ariellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT D19S412."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX Hara S., Kobayashi C., Hamada J., Imura N.;
RT "Characterization of human hypoxia-inducible factor-3 alpha."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007193; AAD22668.1; -.
DR EMBL: AB054067; BAB69689.1; -.
DR InterPro: IPR001092; HLH dim.
DR InterPro: IPR003015; HLH MYC.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR00785; NCTRNSLOCATR.
DR SMART: SM00353; HLH. 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 667 AA; 72404 MW; 67B8794FEF9DCCFAB CRC64;

```

```

Query Match 49.0%; Score 802; DB 4; Length 667;
Best Local Similarity 68.0%; Pred. No. 3.4e-72;
Matches 172; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

QY 9 RSNTELREKSRDARSRSOETEVLYQALHTLPFARGVSAHLDKASIMRLTISYLRMR 68
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 7 RSTTELREKSRDARSRSOETEVLYQALHTLPFARGVSAHLDKASIMRLTISYLRMR 66
QY 69 LCAAGKRGATGRLDPGPGFRHGTNRGRHGLPVGKCO-----QAPGPOSV 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 67 LCAAG-----EMNQYKKEGEPDADCYLKALEGVWVLTAEQDAY 106
QY 118 DLCSSTLIHNPPTGNTS-LELIGHSTFDTHPCDOELDALTPRPMLSKKLEAPTER 176
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 107 -----LSENVSKHLGSQLLELGHSTFDTHPCDOELDALTPRPMLSKKLEAPTER 160
QY 177 HFSLRMKSTLTSRGRITNLKATWKVLCGSHMRAYRPPAOTSPAGSPRSPPLQCLVLI 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 161 HFSLRMKSTLTSRGRITNLKATWKVLCGSHMRAYRPPAOTSPAGSPRSPPLQCLVLI 220
QY 237 CEALPPLPFDGA 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 221 CEALP-----HPGS 229

RESULT 6
09HA12 PRELIMINARY; PRT: 648 AA.
ID 09HA12;
AC 09HA12;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE HYPOTHETICAL 70.0 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO;  
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,  
RA Magatsuna M., Hosori T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,  
RA Niinomiya K., Iwayanagi T.;  
RT "NEO human cDNA sequencing project."  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AK021653; BAB13865.1; -  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR00014; PAS.  
DR Pfam: PF00989; PAS; 2.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
SQ SEQUENCE 648 AA; 69994 MW; EBERC744BC3F148E CRC64;

Query Match 35.0%; Score 573; DB 4; Length 648;  
Best Local Similarity 61.0%; Pred. No. 4; ie-49;  
Matches 125; Conservative 11; Mismatches 27; Indels 42; Gaps 5;

OY 57 MRLTSLYLRMRHLCAGGKGRATGRLPRGPGFRHGRHGRHGLPVGKQ----- 109  
DB 1 MRLTSLYLRMRHLCAGG-----EMNQGAGGEPPLDACYLKALEGF 40  
OY 110 ----QAPGQSDLCSSSLHNPTPGTNE-LELIGHSTDFHPCQDELQDALTPRP 164  
DB 41 VVNLTEGDMAY-----LSENVSKHLGSLQLEIGHSTDFHPCQDELQDALTPRQ 94  
OY 165 LSKKKLEAPRERHPSLRKSTLTSRGRTLNKATWVLCGSHMRAYKPPAOTSPPAGSP 224  
DB 95 LSRKKVAPRERCPSTLSRGRTLNKATWVLCGSHMRAYKPPAOTSPPAGSP 154  
OY 225 RSEPPLOCLVLCGALPQLPFHDA 249  
DB 155 DSEPPLOCLVLCGALP-----HPGS 175

RESULT 7  
O9HAM5 PRELIMINARY; PRT; 199 AA.  
AC O9HAM5;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE HYPOTHETICAL 21.6 KDA PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO;  
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara K., Takeuchi K.,  
RA Arita M., Nabejima T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;  
RT "NEO human cDNA sequencing project."  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AK021421; BAB13819.1; -  
SQ SEQUENCE 199 AA; 21597 MW; 646CB06A09B47C3 CRC64;

Query Match 31.7%; Score 518.5; DB 4; Length 199;  
Best Local Similarity 68.0%; Pred. No. 2; ie-44;  
Matches 104; Conservative 9; Mismatches 13; Indels 27; Gaps 2;

OY 120 CSSSLHNPTPGTNE-----SLEIGHSTDFHPCQDELO 156  
DB 14 CTSMLTRCPSPASAPTWRPLSCASPSACTASAPOLELGHSTFDFHPCQDELO 73  
OY 157 DALTPRPNLSKKLEAPRERHPSLRKSTLTSRGRTLNKATWVLCGSHMRAYKPPA 216  
DB 74 DALTPRPNLSKKLEAPRERHPSLRKSTLTSRGRTLNKATWVLCGSHMRAYKPPA 133  
OY 217 QTSPPAGSPSEPPLOCLVLCGALPQLPFHDA 249  
DB 134 QTSPPAGSPSEPPLOCLVLCGALP-----HPGS 162

RESULT 8  
O9YIB9 PRELIMINARY; PRT; 811 AA.  
AC O9YIB9;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPXIA-INDUCIBLE FACTOR-1 ALPHA.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RA Takahashi T.;  
RT "Molecular cloning and expression of an avian cDNA for hypoxia-inducible factor-1 alpha in embryonic ventricular myocytes."  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB013746; BAA34234.2; -  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR00014; PAS.  
DR Pfam: PF00785; PAC; 1.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
SQ SEQUENCE 811 AA; 90542 MW; D14CD9FC98F064CB CRC64;

Query Match 30.7%; Score 503; DB 13; Length 811;  
Best Local Similarity 49.8%; Pred. No. 6; ie-42;  
Matches 119; Conservative 20; Mismatches 72; Indels 28; Gaps 5;

OY 9 RSNTELRKESRDPAARSQETVELYQLAHTLPFANGVAHLDDKASIMLTSLYLRMR 68  
DB 12 RISSEKREKSRDPAARCRSESEVEFEIHLAQLPLPTVSAHLDDKASIMLTSLYLRMR 71  
OY 69 LCAAGCKGRAT-----GRLLPRGPGGFRHGRHGLPVGKCGQAPGQSDLCSS 123  
DB 72 LLDAGLELETANNEKELNCYLKAIDGFVAVLSEDC-----DNIYMS 113  
OY 124 LHNPTPG-TNFSLEIGHSTDFHPCQDELODALTPRPNLSKKLEAPRERHPSLRM 182  
DB 114 ENNVKCKGGLQF--DLGHGVDFPTHCHDEBELREMLTHRNGVYKKGKGEKNTERSFPLRM 171  
OY 183 KSLTSLRGRTLNKATWVLCGSHMRAYKPPAOTSPPAGSPSEPPLOCLVLCGALP 241  
DB 172 KCLTSLRGRTVNKSAWTKVLCGHITRYDTGNNOTHCG--YKPPMCLVLCGEP 228

RESULT 9  
O9XTA5 PRELIMINARY; PRT; 823 AA.  
AC O9XTA5;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPXIA-INDUCIBLE FACTOR-1 ALPHA.



Query Match	30.3%;	Score 495;	DB 11;	length 258;
Best Local Similarity	44.1%;	Pred. No. 9.4e-42;		
Matches 112;	Conservative 29;	Mismatches 61;	Indels 52;	Gaps 5;

ESULT 12  
96PT9  
Q96PT9  
PRELIMINARY;  
PRT; 735 AA

DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE HYPOXIA-INDUCIBLE FACTOR 1 ALPHA VARIANT.  
GN HIF1A.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP  
RP  
RP TISSUE=LIVER.  
RA Tanaka S., Sugimachi K.:  
RT "Hypoxia-inducible factor-1 alpha variant isolated from human liver  
RT tissue".  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073325; BAB70608.1; -.  
SQ SEQUENCE 735 AA; 82746 MW; 34DD604FB4E418E CRC64;

Query Match	30.3%;	Score 495;	DB 4;	Length 735;
Best Local Similarity	46.2%;	Pred. No. 3.5e-41;		
Matches 114;	Conservative 26;	Mismatches 59;	Indels 48;	Gaps 6

RESULT	13
Q9UB1	
ID	Q9UB1
PRELIMINARY;	
PRT;	836 AA.

DR EMBL: AF050118; AAC68568.1; JOINED.  
DR EMBL: AF050119; AAC68568.1; JOINED.  
DR EMBL: AF050120; AAC68568.1; JOINED.

DR EMBL, AF050122; AAC68568.1; JOINED.  
DR EMBL, AF050123; AAC68568.1; JOINED.  
DR EMBL, AF050124; AAC68568.1; JOINED.  
DR EMBL, AF050125; AAC68568.1; JOINED.  
DR EMBL, AF050126; AAC68568.1; JOINED.  
DR InterPro, IPR001092; HLH\_dgm.  
DR InterPro, IPR003015; HLH\_MYC.  
DR InterPro, IPR001321; Hypoxindf1A.  
DR InterPro, IPR001610; PAC.  
DR InterPro, IPR000014; PAS.  
DR Pfam, PF00785; PAC; 1.  
DR Pfam, PF00989; PAS; 2.  
DR PRINTS, PRO1080; HYPOXIA1F1A.  
DR SMART, SM00353; HLH; 1.  
DR SMART, SM00086; PAC; 1.  
DR SMART, SM00091; PAS; 2.  
DR PROSITE, PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 826 AA; 92636 MW; AB04F198A135C8F CRC64;

Query Match	30.3%;	Score 495;	DB 4;	Length 826;
Best Local Similarity	46.2%;	Pred. No. 4e-41;		
Matches 114;	Conservative 26;	Mismatches 59;	Indels 48;	Gaps 6;

QY	11	NTLELRKESRPAARSRSQETEVLYQLAHLTPRACVSAHLIDKASIMRLTISYLRKHLRC	70
Db	14	SSERRKESRPAARSRRRKRSESEVFETLAHQPLPHNVSSHLDKASVMRLTISYLRKLL	73
QY	71	AAGKRGKATGRLLPEGGGFGHGRHGRHGLPRGKCOQA-----PGQSYDLC	120
Db	74	DAG-----DLDIEDMKKAQMNCFYLKALDGFVWVLT	105
QY	121	SSSLIH-----NPPPG-TNFSLELIGHSIFDEIHPDCODELODALTPRPNLSKKKLEAPT	174
Db	106	DGDMDIYISDNVNYKMGTLQF--ELTGHVSFDEPTPHCDHEMKREMLTHRNGLVKKGQEQWT	163
QY	175	ERHFSLRKSKLTLSNGRFLNLKATWKLKSGHGNKRAKPPRQTPSAGSPRSEPPLOCLV	234
Db	164	QSFELRMRKCVLTISGRTRMNKISATWKLHCGTHIHYDNTNSPOCG--YKPPPTCLV	221
QY	235	LICEAIP 241	
Db	222	LICEPIP 228	
RESULT 14			
ID	Q9WTU9	PRELIMINARY:	PRT: 823 AA.
AC	Q9WTU9:		
DT	01-NOV-1999	(TREMBLrel. 12, Created)	
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	HYOXA-INDUCIBLE FACTOR-1 ALPHA.		
GN	H1FLA.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SPRAGUE DAWLEY; TISSUE=KIDNEY;		
RX	MEDLINE=21417706; PubMed=11526200;		
RA	Zou A.P., Yang Z.Z., Li P.L., Cowley AW J.R.;		
RT	"Oxygen-dependent expression of hypoxia-inducible factor-1alpha in		
RL	renal medullary cells of rats."		
RL	Physiol. Genomics (Online) 6:159-168(2001).		
DR	EMBL; AF057308; AAD24413.1;--		
DR	InterPro: IPR001092; HLH_dlm.		
DR	InterPro: IPR003015; HLH_Myc.		
DR	InterPro: IPR001321; Hypoxindf1A.		
DR	InterPro: IPR000104; PAS.		
DR	Pfam: PF00785; PAC; 1.		
DR	Pfam: PF00989; PAS; 2.		
DR	PRINTS; PRO1080; HYPOXIA1FLA.		
DR	SMART; SM00353; HLH; 1.		
DR	SMART; SM00086; PAC; 1.		
DR	SMART; SM00091; PAS; 2.		
DR	PROSITE; PS00036; HELIX_LOOP_HELIX; UNKNOWN_1.		
SO	SEQUENCE 823 AA; 92116 MW; ODC907B530766FIF CRC64;		
Query Match 30.0%; Score 491; DB 11; Length 823;			
Best Local Similarity 45.4%; Pred. No. 1e-40; Mismatches 61; Indels 48; Gaps 56;			
Matches 113; Conservative 27; Mismatches 61; Indels 48; Gaps 56;			
QY	9	RSNFLRKRKESRDARSRSQETEVLYQLAHLTPRACVSAHLIDKASIMRLTISYLRMR	68
Db	13	RMSRKRKRKESRDARSRSKSEVFETLAHQPLPHNVSSHLDKASVMRLTISYLRVAK	72
QY	69	LCAAGKRGKATGRLLPEGGGFGHGRHGRHGLPRVKGCOQA-----PGQSYD	118
Db	73	LLGAG-----DLDIEDMKKAQMNCFYLKALDGFVWVLT	104
QY	119	LCSSSLIH-----NPPPG-TNFSLELIGHSIFDEIHPDCODELODALTPRPNLSKKKLEA	172
Db	105	TDDGMDIYISDNVNYKMGTLQF--ELTGHVSFDEPTPHCDHEMKREMLTHRNGPVRKQKQ	162

QY	173	PTERHSLRMSTSTLRRTINLKAATMKVHCGSHMRAYKPRQGTSPASPREPELOC	23
Db	163	NTORSEFLRMCTLTLSRRTINLNKSAITWKVLDCHGTHIVDTSSNOPQCG--YKPPMTC	220
QY	233	LVLICEAIP 241	
Db	221	LVLICEPIP 229	
RESULT	15		
ID	Q9JHS1	PRELIMINARY; PRT; 874 AA.	
AC	Q9JHS1		
DT	01-OCT-2000 (TREMblrel, 15, Created)		
DT	01-OCT-2000 (TREMblrel, 15, Last sequence update)		
DT	01-DEC-2001 (TREMblrel, 19, Last annotation update)		
DE	HYPOXIA INDUCIBLE FACTOR 2 ALPHA.		
GN	HIF-2A.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
RX	NCBI_TaxID=10116;		
RY	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER.		
RX	MEDLINE=21134367; PubMed=11237857;		
RA	Kietzmann T., Cornesse Y., Brechtel K., Modarresi S., Jungemann K.;		
RT	"Particulous expression of the mRNA of the three hypoxia-inducible		
RL	factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";		
RT	Biochem. J. 354:531-537(2001).		
DR	EMBL: AJ277828; CAB96612.1; "		
DR	InterPro: IPR001092; HLM.dlm.		
DR	InterPro: IPR003015; HLM.Myc.		
DR	InterPro: IPR001067; NucleinsLocator.		
DR	InterPro: IPR001610; PAC.		
DR	InterPro: IPR000014; PAS.		
DR	Pfam: PF00785; PAC; 1.		
DR	Pfam: PF00989; PAS; 2.		
DR	PRINTS: PRO0785; NCTRNSLOCATR.		
DR	SMART: SM00353; HLH; 1.		
DR	SMART: SM00086; PAC; 1.		
DR	SMART: SM00091; PAS; 2.		
DR	PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.		
SO	SEQUENCE 874 AA; 96718 MW; ALF080B24369796 CRC64;		

```

Query Match          29.5%; Score 482; DB 11; Length 874;
Best Local Similarity 42.2%; Pred. No. 8,9e-40;
Matches 125; Conservative 39; Pident 96; Indels 36; Gaps 11.

QY      1 MALGQVRVSNTELKERSRDAARSRQSETVLVOLAHTELRFANGVSHLDKASIMRLT 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 MPADEKRRSSSELKERSRDAARSRKSETVFELAEHLRLPHSVSSHLDKASIMRLA 60
QY      61 ISYLRMHLR---CAAGCKRGKRGATRLRLDEGGGFRHGTNRGRHGLPYGKCOQAPPOS 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61 ISFLTHKLSSVCSGENESEAADDOM-----DNLTKALGEGF-----TAVTQD 105
QY      117 VDLCSSSLINHPPTNFSLELIGHISDFIHPDQOELQDALTPRP--PNLSKKKLAEPT 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      106 GDMILSEINISFMGLT-QVELTGHISIFPTHPCHDEHLEIRMLTLKTSGGFPKKKNDRST 164
QY      175 ERHFSLRMKSTLTSRGRTLLNLKAATWTKVLHCSGHNRAVYKR-PAQTSPPAGSPREPPQL 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      165 EEDDFMRMKCTYTNNGRTVNLKSAITWTKVLHCGQVRYVYNNCPRHSSLGG--YKEPRLSSL 222
QY      234 VLICEAIPQLPFDHAGATLGLP-QEKTPIS-----TLFPRMLALCLVYKRWPFVQVL 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      223 IIMCPT-QHPSH-----MDIPLDSTKPTLSRHSMDKKFTYCDDRILELVGYHDEELL 273

```

Search completed: September 8, 2002, 07:41:25  
Job time: 443 sec

---